Unlock the Value in your Big Data Reservoir using Oracle Big Data Discovery and Oracle Big Data Spatial and Graph

Robin Moffatt, Head of R&D (Europe), Rittman Mead

DW & Big Data Global Leaders Program, June 2016 - Oslo, Norway

info@rittmanmead.com www.rittmanmead.com @rittmanmead

rittmanmead

About Me

- Head of R&D (Europe), Rittman Mead
  - Previously OBIEE/DW developer at large UK retailer
  - Previously SQL Server DBA, Business Objects, DB2, COBOL….
- Oracle ACE
- Frequent blogger : http://ritt.md/rmoff
- Twitter: @rmoff
- IRC: rmoff / #obihackers / freenode

About Rittman Mead

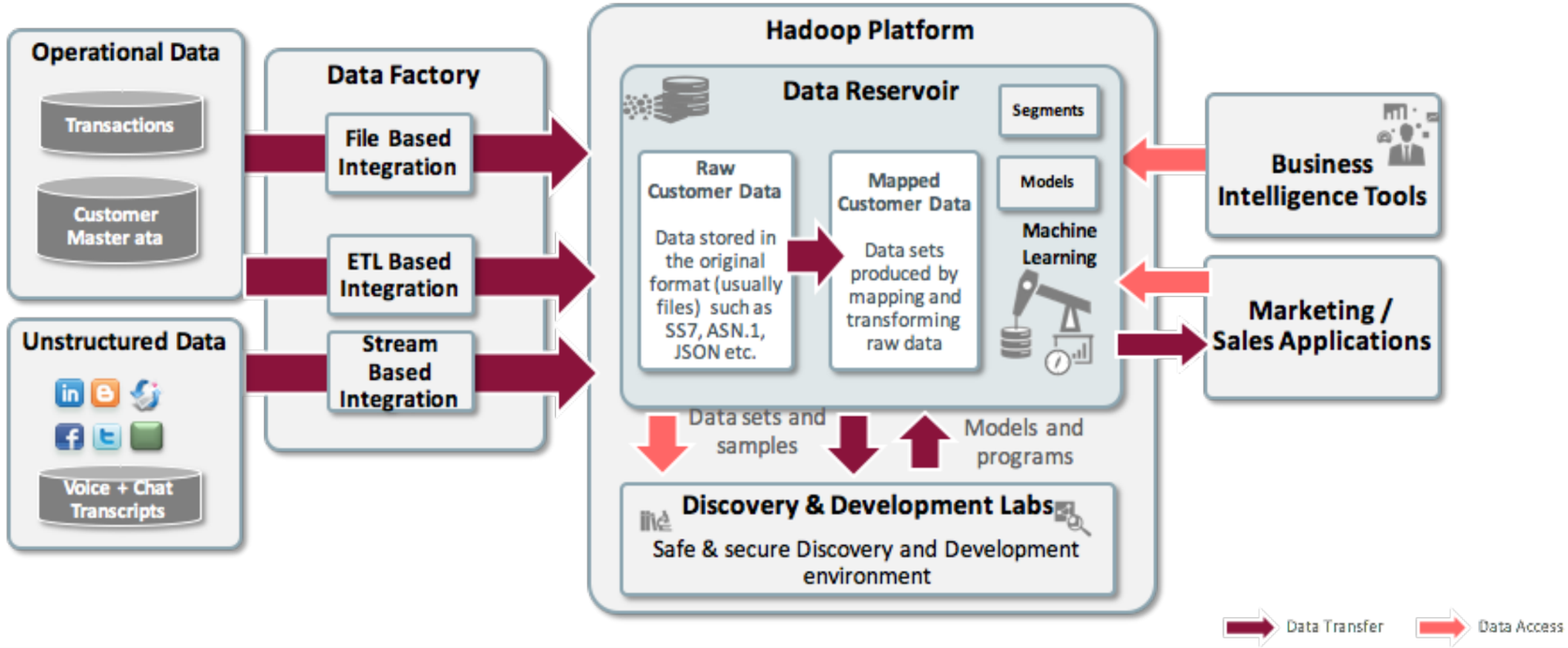
- Oracle Gold Partner with offices in the UK and USA (Atlanta)
- 70+ staff delivering Oracle BI, DW, Big Data and Advanced Analytics projects
- 2 Oracle ACE Directors + 2 Oracle ACEs
- Significant web presence with the Rittman Mead Blog (http://www.rittmanmead.com)
- Regular sers of social media (Facebook, Twitter, Slideshare etc)
- Regular column in Oracle Magazine and other publications
- Hadoop R&D lab for “dogfooding” solutions developed for customers

Many Organisations are Running Big Data Initiatives

- Many customers and organisations are now running initiatives around “big data”
- Some are IT-led and are looking for cost-savings around data warehouse storage + ETL
- Others are “skunkworks” projects in the marketing department that are now scaling-up
- Projects now emerging from pilot exercises
- And design patterns starting to emerge

Common Big Data Design Pattern : “Data Reservoir”

- Typical implementation of Hadoop and big data in an analytic context is the “data lake”
- Additional data storage platform with cheap storage, flexible schema support + compute
- Data lands in the data lake or reservoir in raw form, then minimally processed
- Data then accessed directly by “data scientists”, or processed further into DW



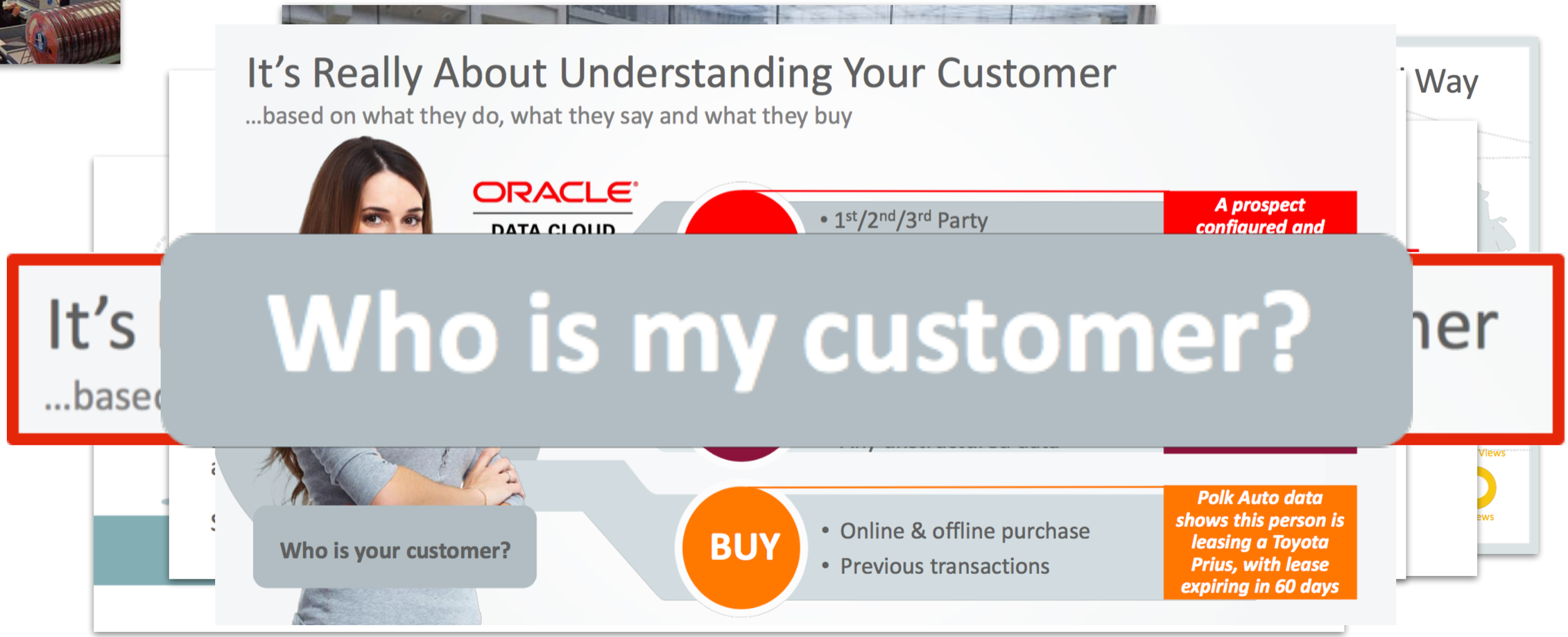
So What is a Data Reservoir?

What Does it Do?

And Does it Replace My Data Warehouse?

An Interesting Question.

Meanwhile, back in the real world…


It's Really About Understanding Your Customer
...based on what they do, what they say and what they buy
ORACLE
DATA CLOUD
• 1st/2nd/3rd Party
A prospect configured and
Who is your customer?
BUY
• Online & offline purchase
• Previous transactions
Polk Auto data shows this person is leasing a Toyota Prius, with lease expiring in 60 days
Who is my customer?

Who is my customer?

Behaviors
Activities
Sentiment
Preferences

More Complete Data About Your Customers

More Attributes and Activities Stored at Scale

True 360° Customer Profile

Connect disparate data

Targeted, personalized customer treatment

Advanced analytics and machine learning

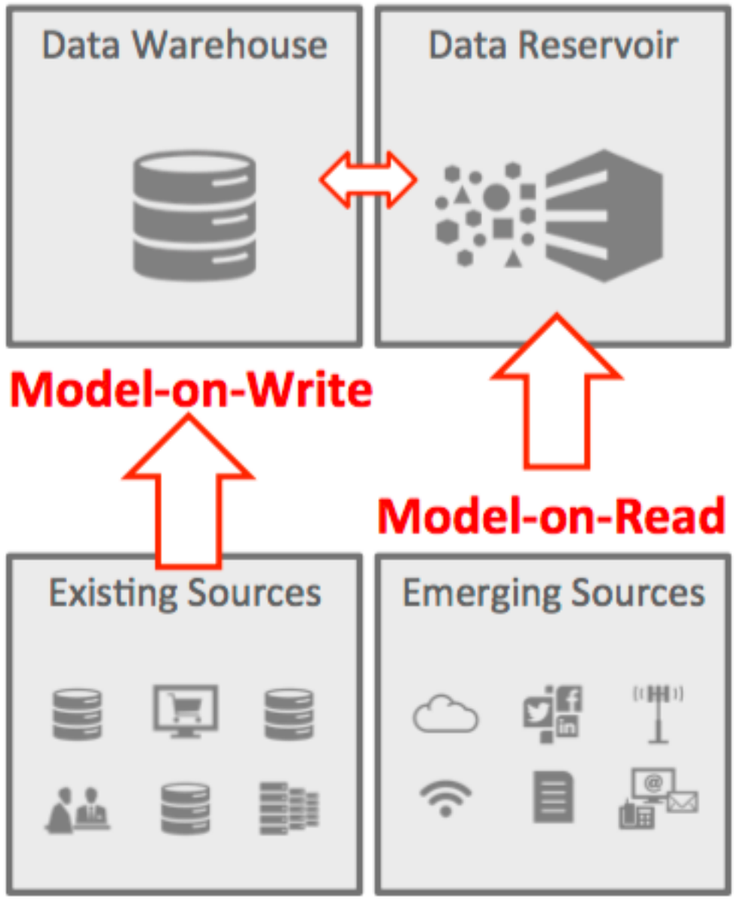
Customer 360-Degree Insight

Data from Real-Time, Social & Internet Sources is *Strange*

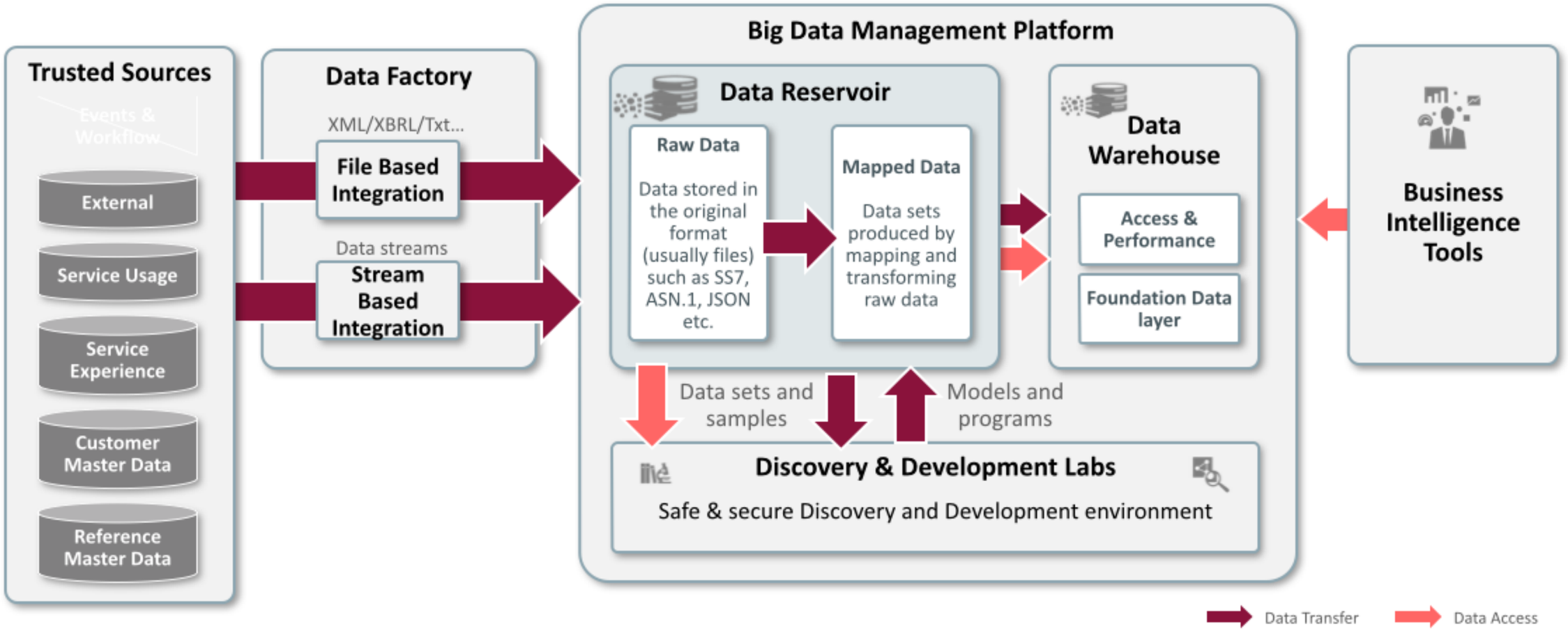
- Typically comes in non-tabular form
  - JSON, log files, key/value pairs
- Users often want it speculatively
  - Haven't though through final purpose
- Schema can change over time
  - Or maybe there isn't even one
- But the end-users want it now
  - Not when your ETL team are next free

Introducing Hadoop - Cheap, Flexible Storage + Compute

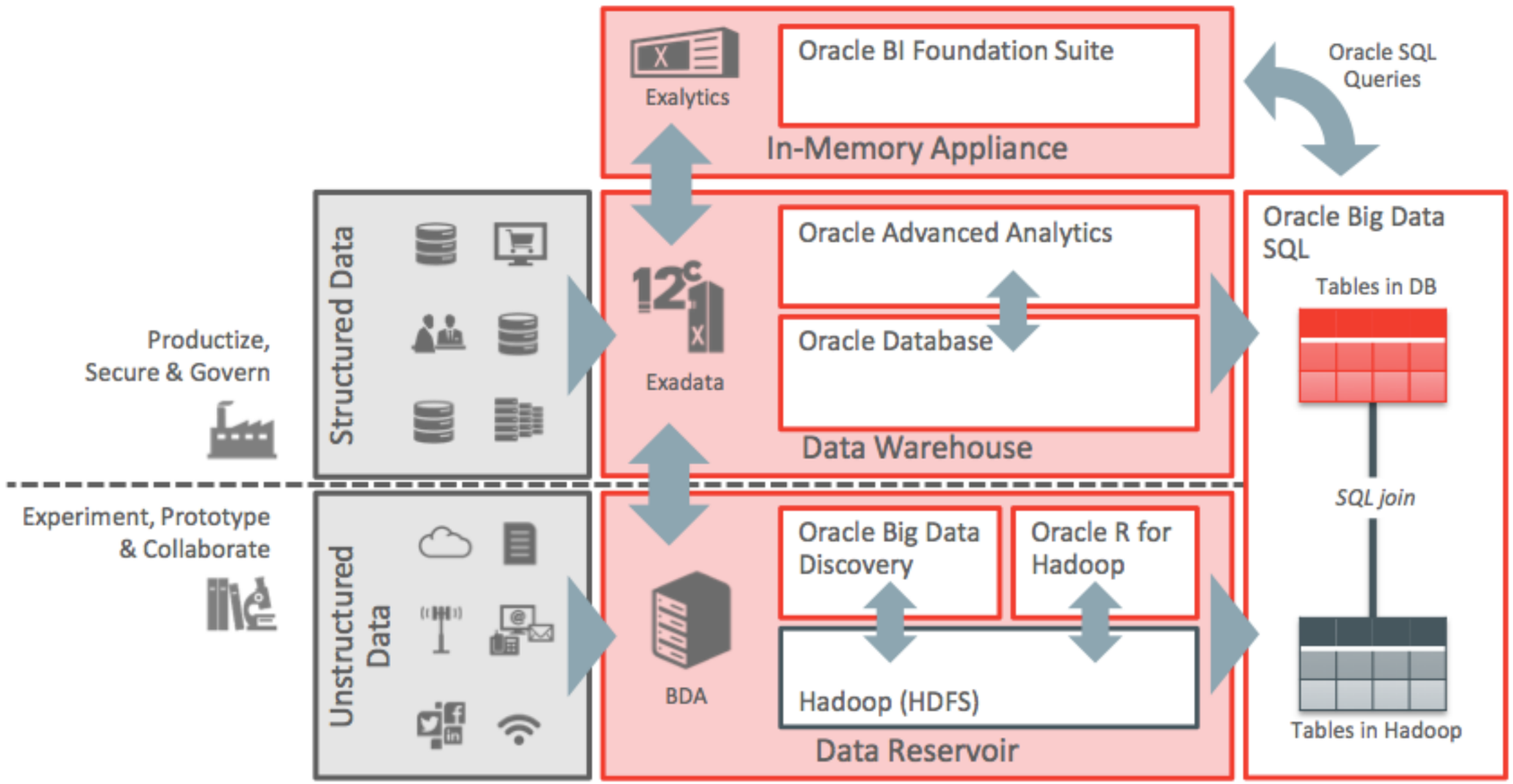
- Hadoop & NoSQL better suited to exploratory analysis of new data
  - Flexible schema - applied by user rather than ETL
  - Cheap expandable storage for detail-level data
  - Better native support for machine-learning and data discovery tools and processes
- Potentially a great fit for our new and emerging customer 360 datasets, and great platform for analysis



Combine with DW for Big Data Management Platform



The Oracle BI, DW and Big Data Product Architecture



- Experiment, Prototype, Collaborate
  - Quickly find, explore, transform, discover and share in BDD
  - Publish results to HDFS
  - Use to build predictive models with Oracle R for Hadoop
- Productize, Secure, Govern
  - Connect published HDFS files to secure Oracle DB using Oracle Big Data SQL
  - No data movement required
  - Seamlessly extends existing DWH and BI investments with non-traditional data in Hadoop
- Available as Engineered Systems

But … These Data Sources are *Strange*

- Typically comes in non-tabular form
  - JSON, log files, key/value pairs
- Users often want it speculatively
  - Haven't though through final purpose
- Schema can change over time
  - Or maybe there isn't even one
- But the end-users want it now
  - Not when your ETL team are next free

But … These Data Sources are *Strange*

collects new and existing data as raw materials

provides broad access to a trusted community

offers a ‘do anything’ sandbox environment

enables agile experimentation, ability to fail fast

Introducing the “Data Lab” for Raw/Unstructured Data

collects new and existing data as raw materials

provides broad access to a trusted community

offers a ‘do anything’ sandbox environment

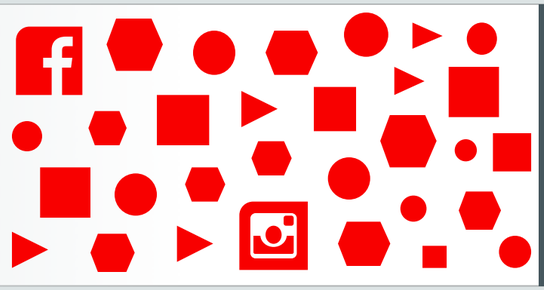
enables agile experimentation, ability to fail fast

Turning Raw Data into Information and Value is Hard

- Specialist skills typically needed to ingest and understand data coming into Hadoop
- Data loaded into the reservoir needs preparation and curation before presenting to users
- How do we staff and scale projects as our use of big data matures?
- *But we've heard a similar story before, a few years ago…*

Data Uncertainty

- Not familiar and overwhelming
- Potential value not obvious
- Requires significant manipulation

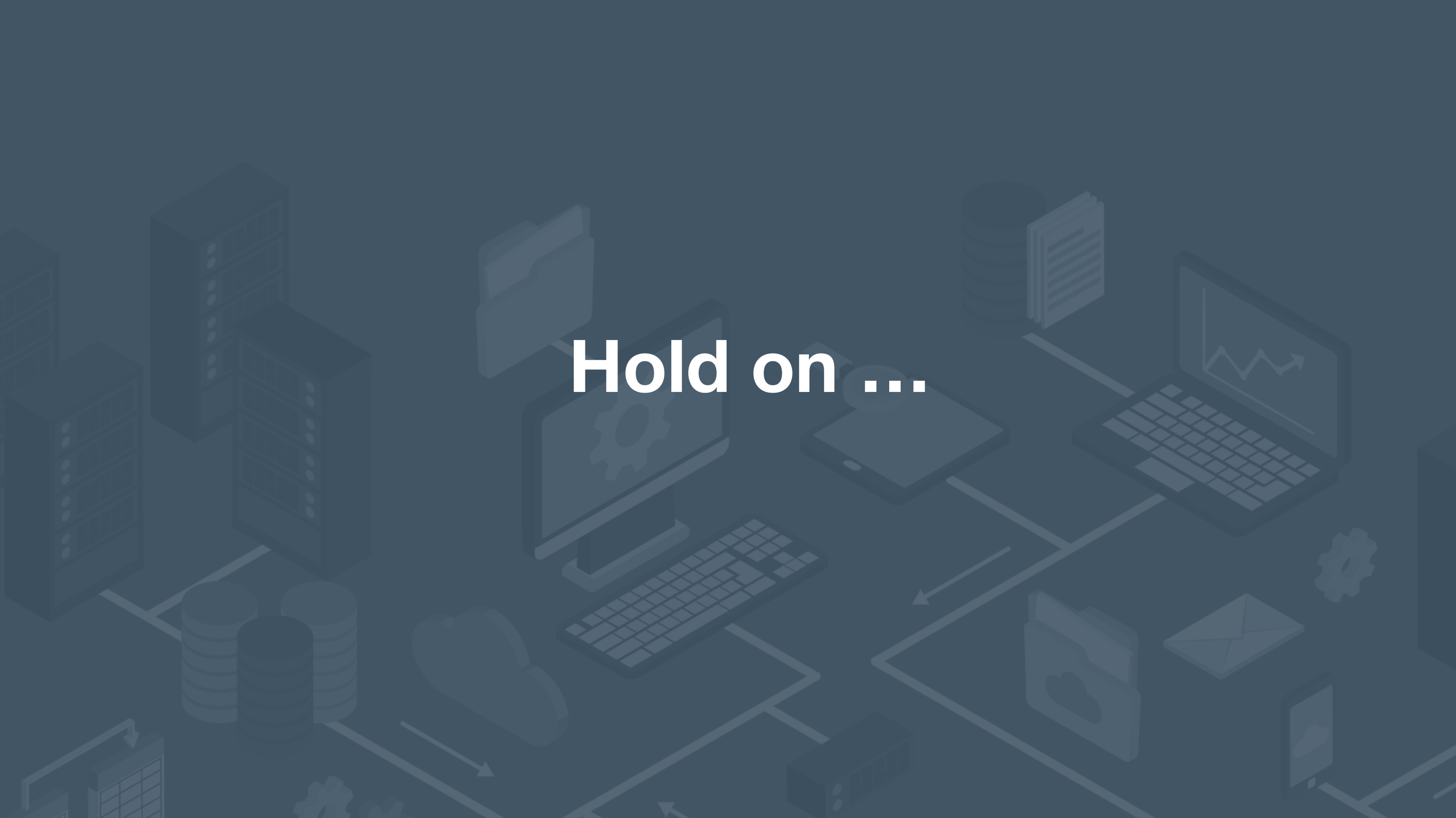


80% effort typically spent on evaluating and preparing data

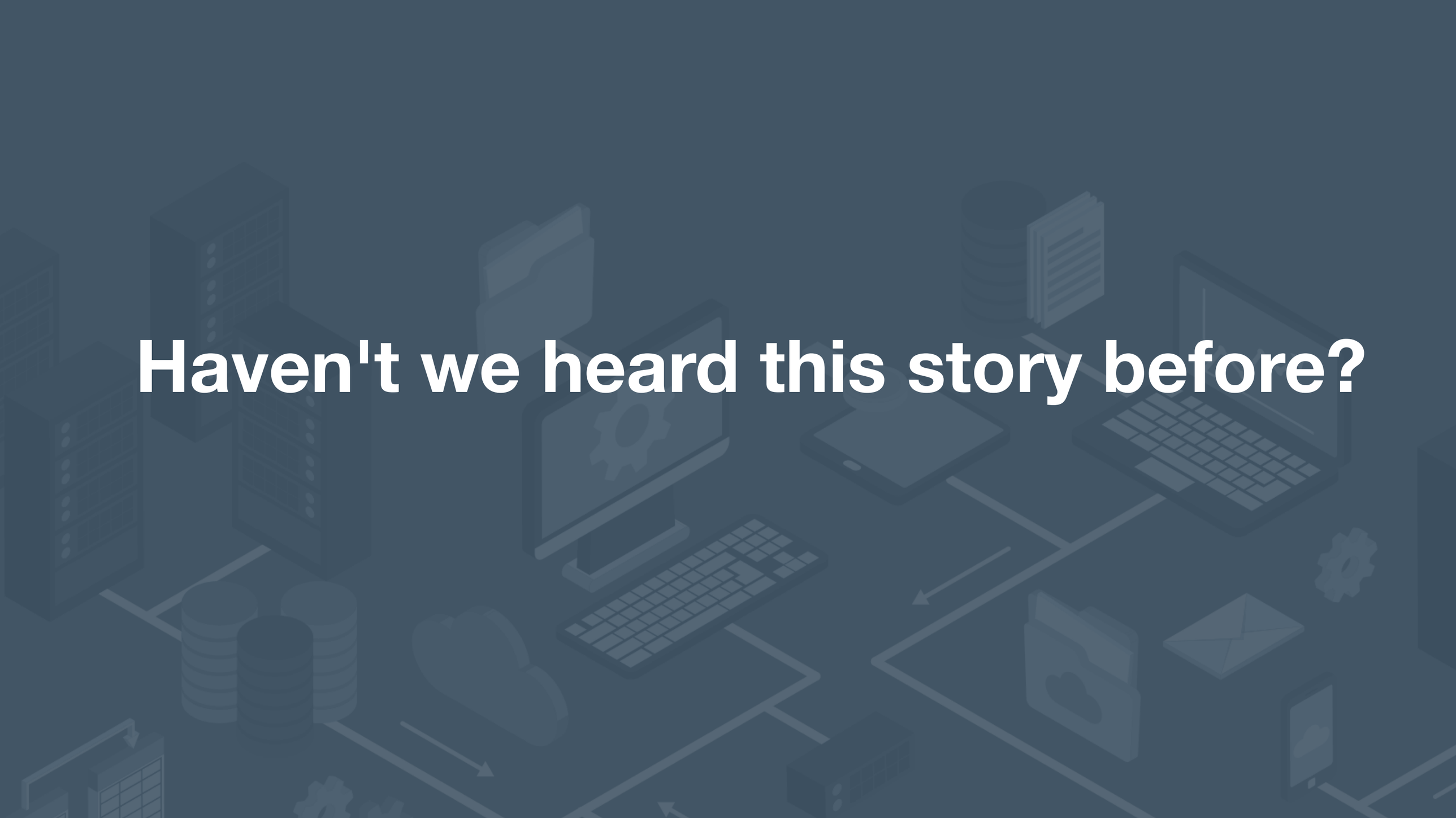
Tool Complexity

- Early Hadoop tools only for experts
- Existing BI tools not designed for Hadoop
- Emerging solutions lack broad capabilities

Overly dependent on scarce and highly skilled resources


Hold on ...

Haven't we heard this story before?



Olympic Games 2012
The Northerner

Yorkshire finishes twelfth in medals table at Olympics

Tourism agency publishes a Yorkshire Heroes map and we can expect a hosting bid from Wetwang in 2020

Martin Wainwright

Monday 13 August 2012
06.44 BST

Comments
41

Save for later

The place is full of them. Photograph: Alamy

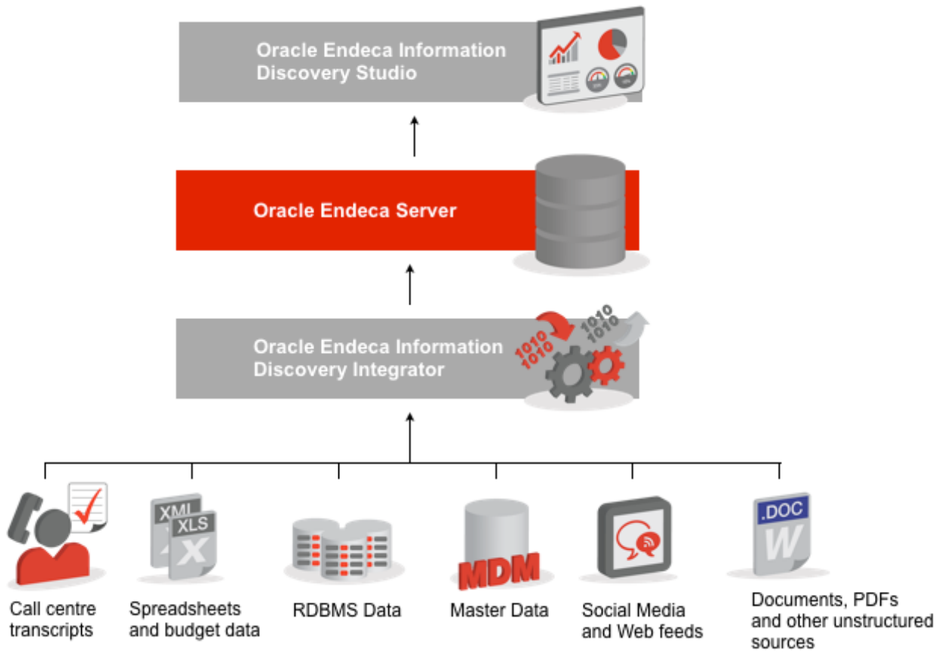
Yorkshire has finished up with seven gold medals, two silver and three bronzes, placing it twelfth in the medal table if regarded as an independent country, as it should be.

What Was Oracle Endeca Information Discovery?

- Part of the acquisition of Endeca back in 2012 by Oracle Corporation
- Based on search technology and concept of “faceted search”
- Data stored in flexible NoSQL-style in-memory database called “Endeca Server”
- Added aggregation, text analytics and text enrichment features for “data discovery”
  - Explore data in raw form, loose connections, navigate via search rather than hierarchies
  - Useful to find out what is relevant and valuable in a dataset before formal modeling

Endeca Server Technology Combined Search + Analytics

- Proprietary database engine focused on search and analytics
- Data organized as records, made up of attributes stored as key/value pairs
- No over-arching schema,
  no tables, self-describing attributes
- Endeca Server hallmarks:
  - Minimal upfront design
  - Support for “jagged” data
  - Administered via web service calls
  - “No data left behind”
  - “Load and Go”
- But … limited in scale (>1m records)
  - *… what if it could be rebuilt on Hadoop?*



2016

Press Release

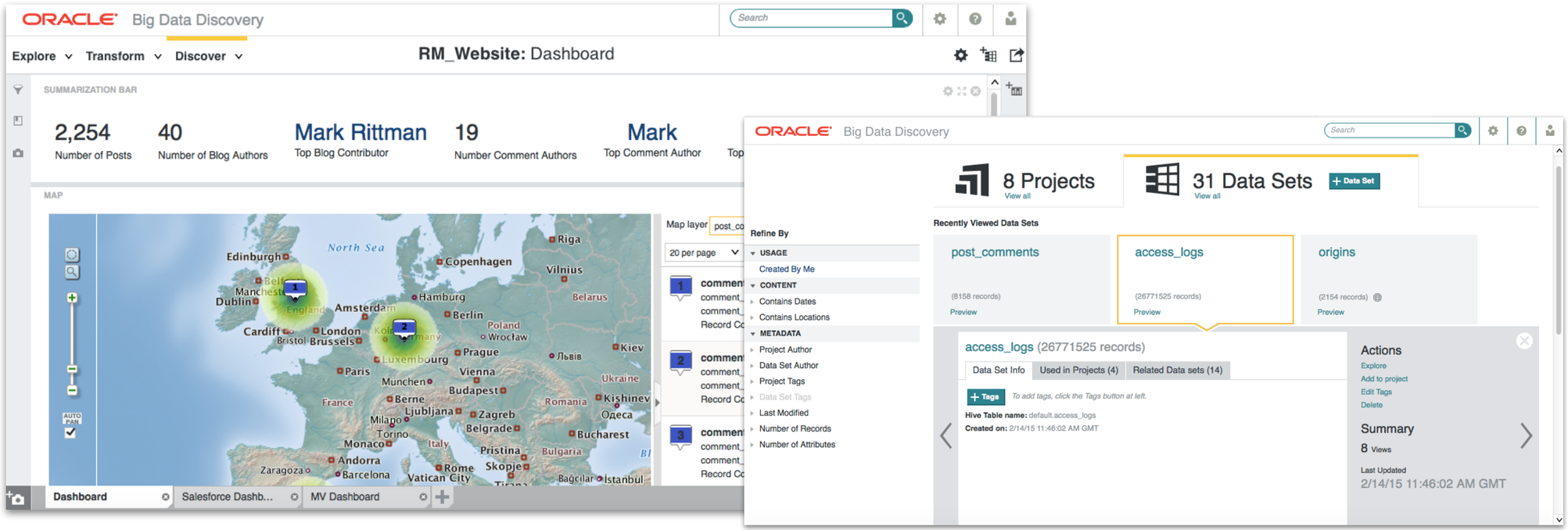
Oracle Advances Vision for Enterprise Big Data

New technologies help integrate big data with enterprise computing, making big data insights a reality

Strata + Hadoop World, San Jose, Calif.—Feb 19, 2015

Oracle Big Data Discovery

- A visual front-end to the Hadoop data reservoir, providing end-user access to datasets
- Catalog, profile, analyse and combine schema-on-read datasets across the Hadoop cluster
- Visualize and search datasets to gain insights, potentially load in summary form into DW



What Does Big Data Discovery Do?

- Provide a visual catalog and search function across data in the data reservoir
- Profile and understand data, relationships, data quality issues
- Apply simple changes, enrichment to incoming data
- Visualize datasets including combinations (joins)

Data Lake Navigator

Top Notch Data Wrangler

Big Data Dashboards

Example Scenario : Social Media Analysis

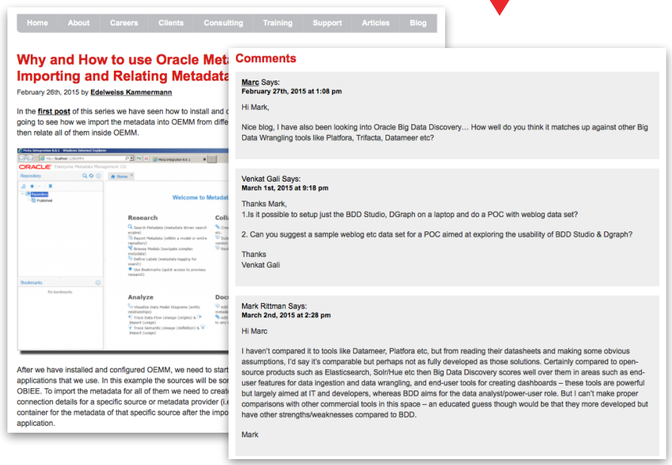
- Rittman Mead want to understand drivers and audience for their website
  - What is our most popular content? Who are the most in-demand blog authors?
  - Who are the influencers? What do they read?
- Three data sources in scope:

RM Website Logs

```
202.46.48.28 - - [23/Feb/2014:01:03:37 +0000] "GET / HTTP/1.1" 200
40440 "-" "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.0)"
119.63.193.195 - - [23/Feb/2014:01:04:57 +0000] "GET / HTTP/1.1" 200
40440 "-" "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.0)"
66.249.74.116 - - [23/Feb/2014:01:26:31 +0000] "GET /2011/03/update-
on-the-oracle-press-book/ HTTP/1.1" 200 11585 "-" "Mozilla/5.0
(compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
66.249.74.116 - - [23/Feb/2014:01:52:48 +0000] "GET /blog/page/336/
HTTP/1.1" 200 10600 "-" "Mozilla/5.0 (compatible; Googlebot/2.1;
+http://www.google.com/bot.html)"
180.76.5.222 - - [23/Feb/2014:01:56:18 +0000] "GET /2010/09/rittman-
mead-at-oracle-openworld-2010-san-francisco/ HTTP/1.1" 304 - "-"
"Mozilla/5.0 (compatible; Baiduspider/2.0; +http://www.baidu.com/
search/spider.html)"
180.76.5.166 - - [23/Feb/2014:02:33:53 +0000] "GET /2009/05/dynamic-
sql/ HTTP/1.1" 200 11978 "-" "Mozilla/5.0 (compatible; Baiduspider/
2.0; +http://www.baidu.com/search/spider.html)"
91.232.96.2 - - [23/Feb/2014:02:37:19 +0000] "GET /blog/page/33/
HTTP/1.1" 200 95831 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows
NT 5.1; SV1)"
91.232.96.2 - - [23/Feb/2014:02:38:01 +0000] "GET /2012/05/
hfm-11-1-2-2-%E2%80%93-new-features-part-1/ HTTP/1.1" 200 50539 "-"
"Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)"
91.232.96.2 - - [23/Feb/2014:02:38:09 +0000] "POST /wp-comments-
post.php HTTP/1.1" 302 - "http://31.221.34.123/2012/05/hfm-11-1-2-2-
%E2%80%93-new-features-part-1/" "Mozilla/4.0 (compatible; MSIE 6.0;
Windows NT 5.1; SV1)"
91.232.96.2 - - [23/Feb/2014:02:38:10 +0000] "GET /2012/05/
```

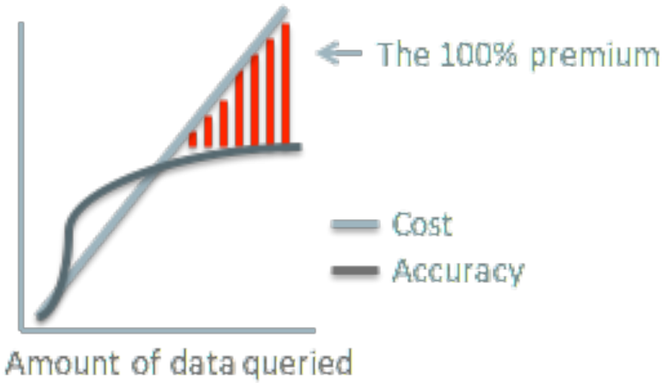
Twitter Stream

Website Posts, Comments etc



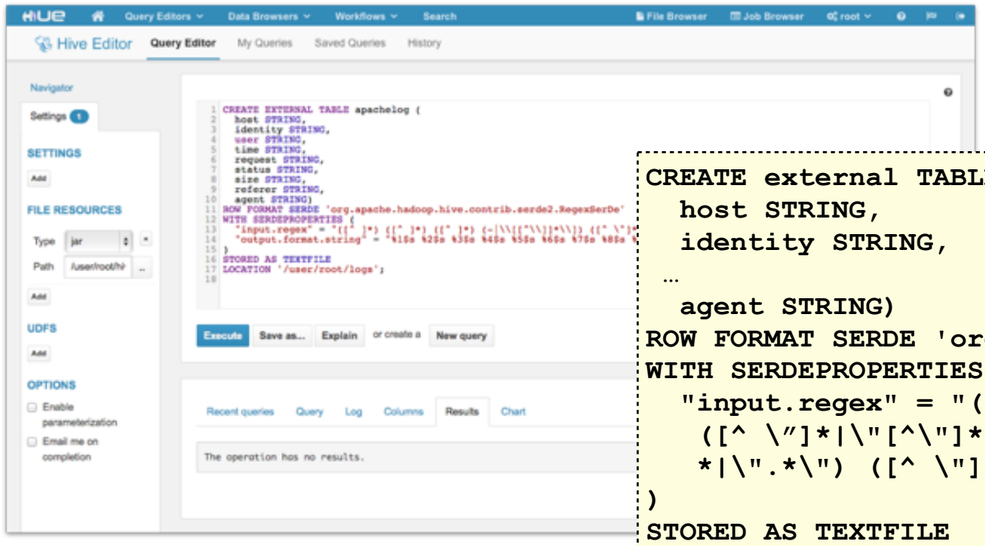
Ingesting Data to Big Data Discovery

- Data has to be ingested into DGraph engine before analysis, transformation
- Primary route is from existing data on HDFS, exposed through Hive
  - Can either define an automatic Hive table detector process, or manually trigger
- Option also to import data from flat file or JDBC
  - Uses HDFS to store it
- Typically ingests 1m row random sample
  - 1m row sample provides > 99% confidence that answer is within 2% of value shown no matter how big the full dataset (1m, 1b, 1q+)
  - Makes interactivity cheap - representative dataset



Enabling Raw Data for Access by Big Data Discovery

- Relies on datasets in Hadoop being registered with Hive Catalog
- Presents semi-structured and other datasets as tables, columns
- Hive SerDe and Storage Handler technologies allow Hive to run over most datasets
- Hive tables need to be defined before dataset can be used by BDD




```
CREATE external TABLE apachelog_parsed(
  host STRING,
  identity STRING,
 …
  agent STRING)
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.RegexSerDe'
WITH SERDEPROPERTIES (
  "input.regex" = "([^]*) ([^]*) ([^]*) (-|\\[^\\]*\\])
   ([^ \"]*|\"[^\"]*\")(-|[0-9]*) (-|[0-9]*)(?: ([^ \"]
   *|\".*\") ([^ \"]*|\".*\"))?"
)
STORED AS TEXTFILE
LOCATION '/user/flume/rm_website_logs;
```

Ingesting Logs and Tweet Data Samples into DGraph

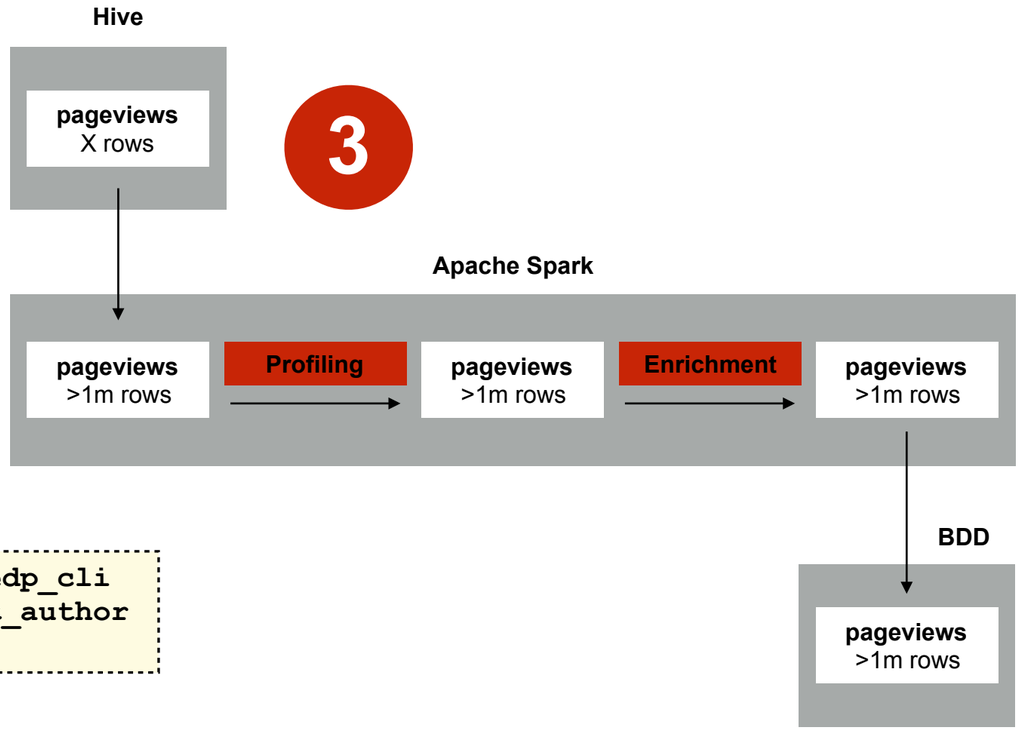
- Tweets and Website Log Activity stored already in data reservoir as Hive tables
- Upload triggered by manual call to BDD Data Processing CLI
  - Runs Oozie job in the background to profile, enrich and then ingest data into DGraph

2

```
{
  "@class" : "com.oracle.endeca.pdi.client.config.workflow.
              ProvisionDataSetFromHiveConfig",
  "hiveTableName" : "rm_linked_tweets",
  "hiveDatabaseName" : "default",
  "newCollectionName" : "edp_cli_edp_a5dbdb38-b065…",
  "runEnrichment" : true,
  "maxRecordsForNewDataSet" : 1000000,
  "languageOverride" : "unknown"
}
```

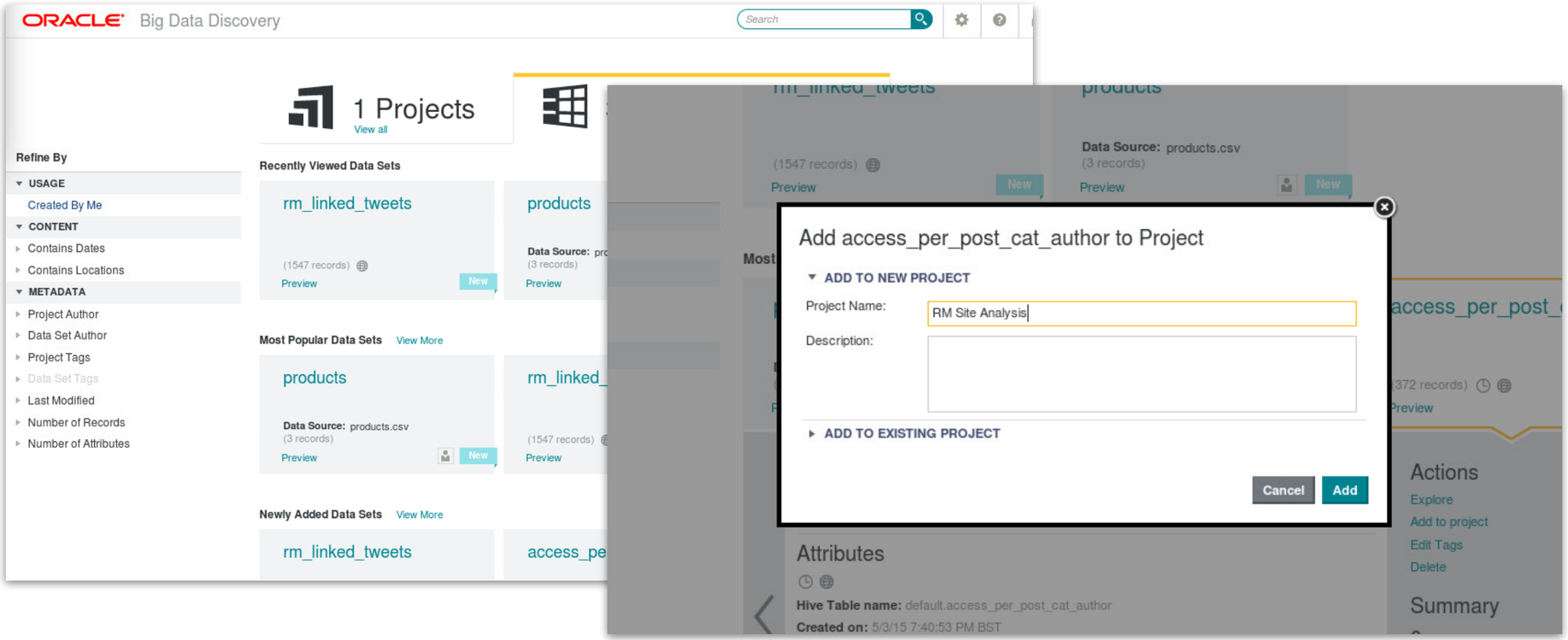
1

```
[oracle@bddnode1 ~]$ cd /home/oracle/Middleware/BDD1.0/dataprocessing/edp_cli
[oracle@bddnode1 edp_cli]$ ./data_processing_CLI -t access_per_post_cat_author
[oracle@bddnode1 edp_cli]$ ./data_processing_CLI -t rm_linked_tweets
```



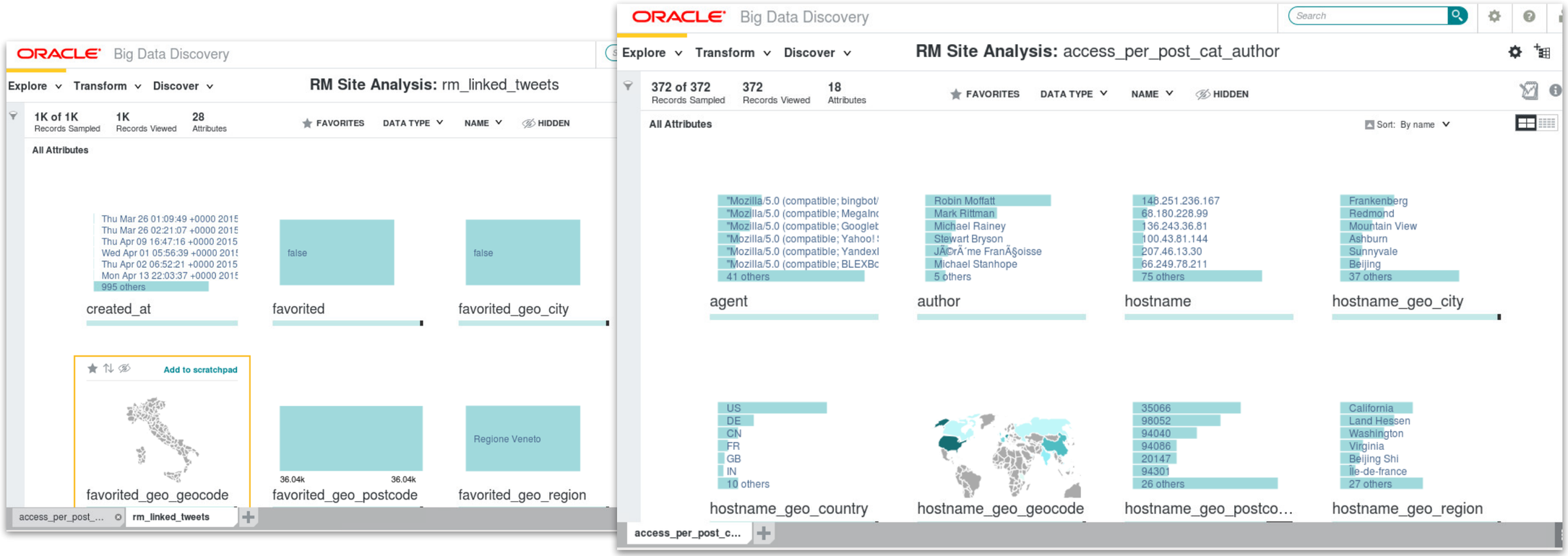
View Ingested Datasets, Create New Project

- Ingested datasets are now visible in Big Data Discovery Studio
- Create new project from first dataset, then add second



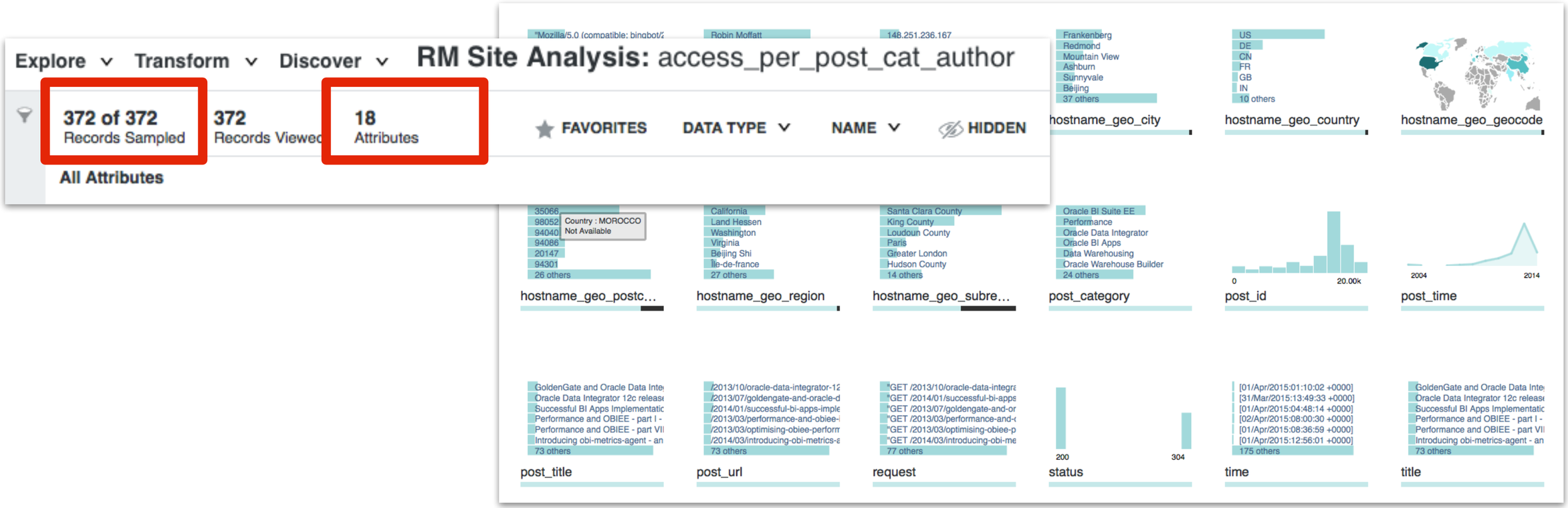
Automatic Enrichment of Ingested Datasets

- Ingestion process has automatically geo-coded host IP addresses
- Other automatic enrichments run after initial discovery step, based on datatypes, content



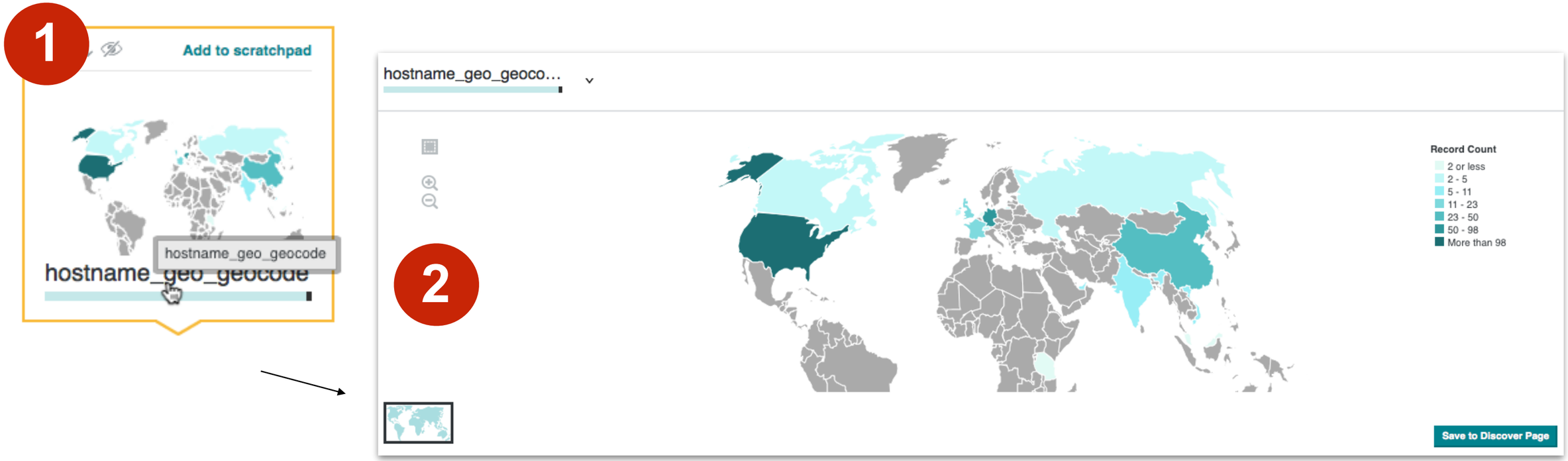
Initial Data Exploration On Uploaded Dataset Attributes

- For the ACCESS_PER_POST_CAT_AUTHORS dataset, 18 attributes now available
- Combination of original attributes, and derived attributes added by enrichment process



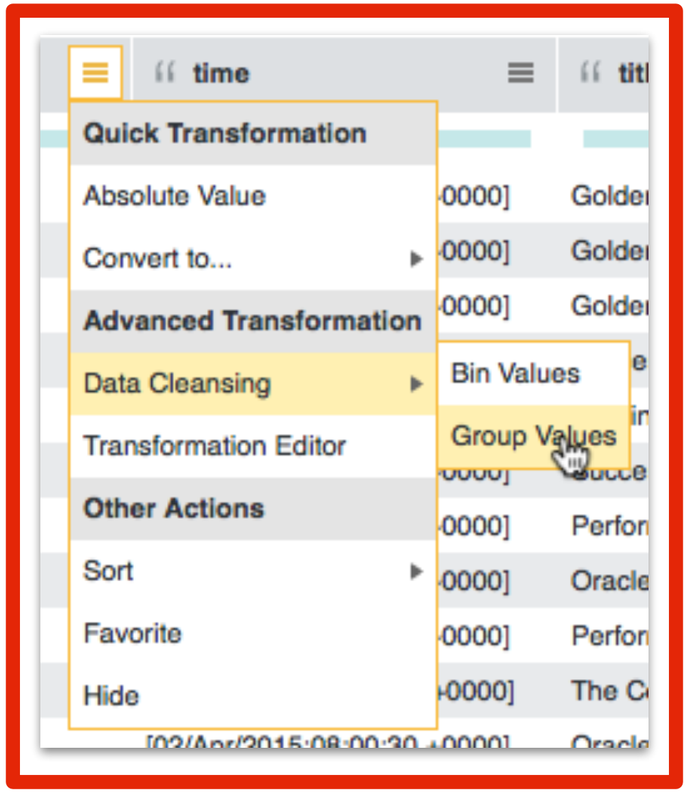
Explore Attribute Values, Distribution using Scratchpad

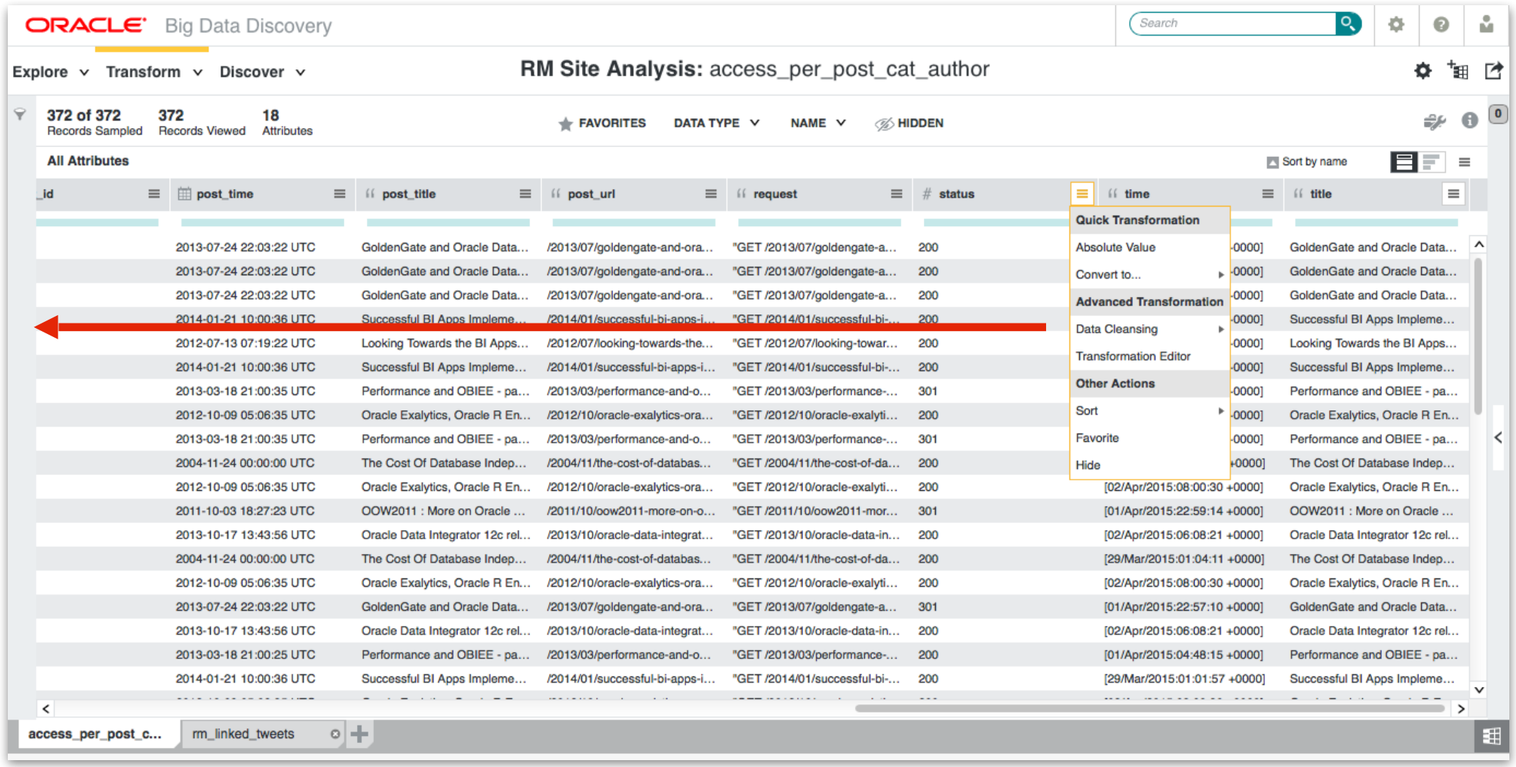
- Click on individual attributes to view more details about them
- Add to scratchpad, automatically selects most relevant data visualisation



Data Transformation & Enrichment

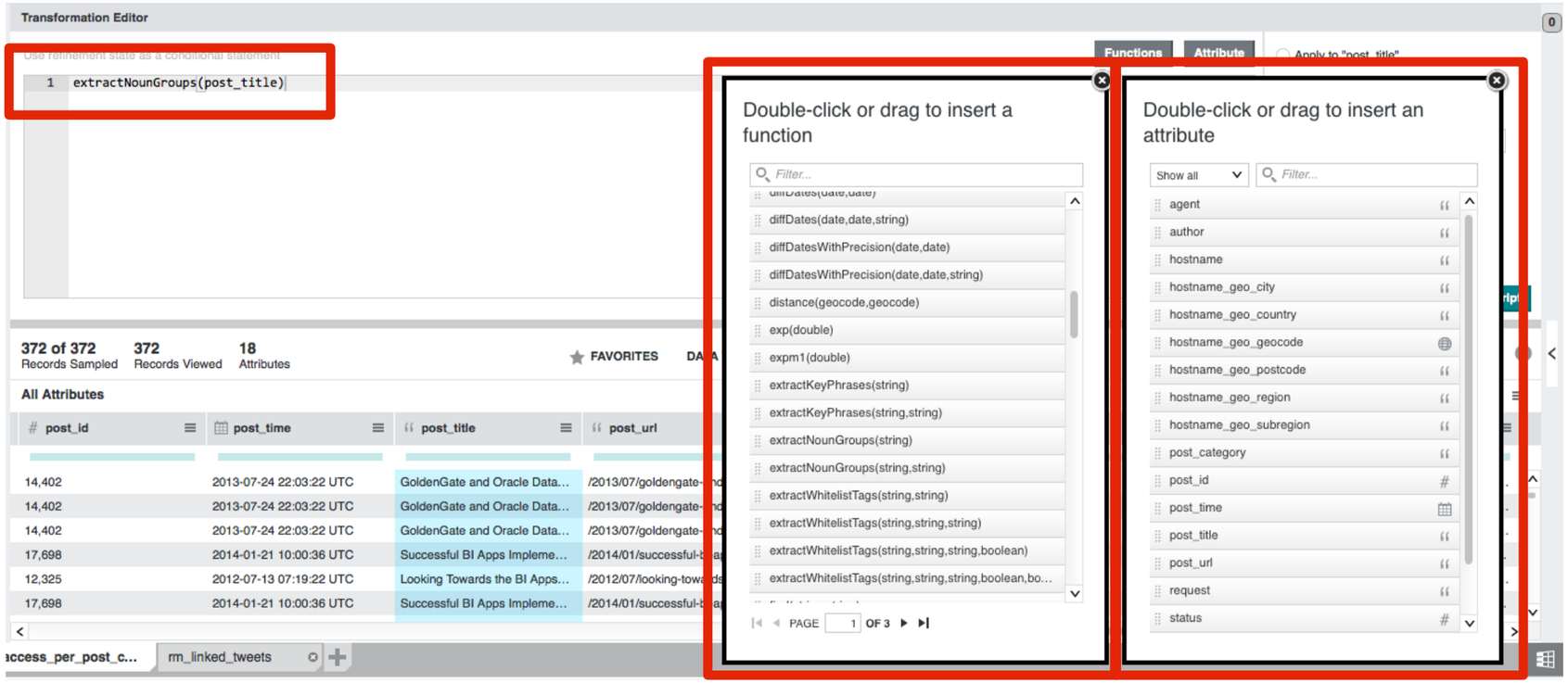
- Data ingest process automatically applies some enrichments - geocoding etc
- Can apply others from Transformation page - simple transformations & Groovy expressions





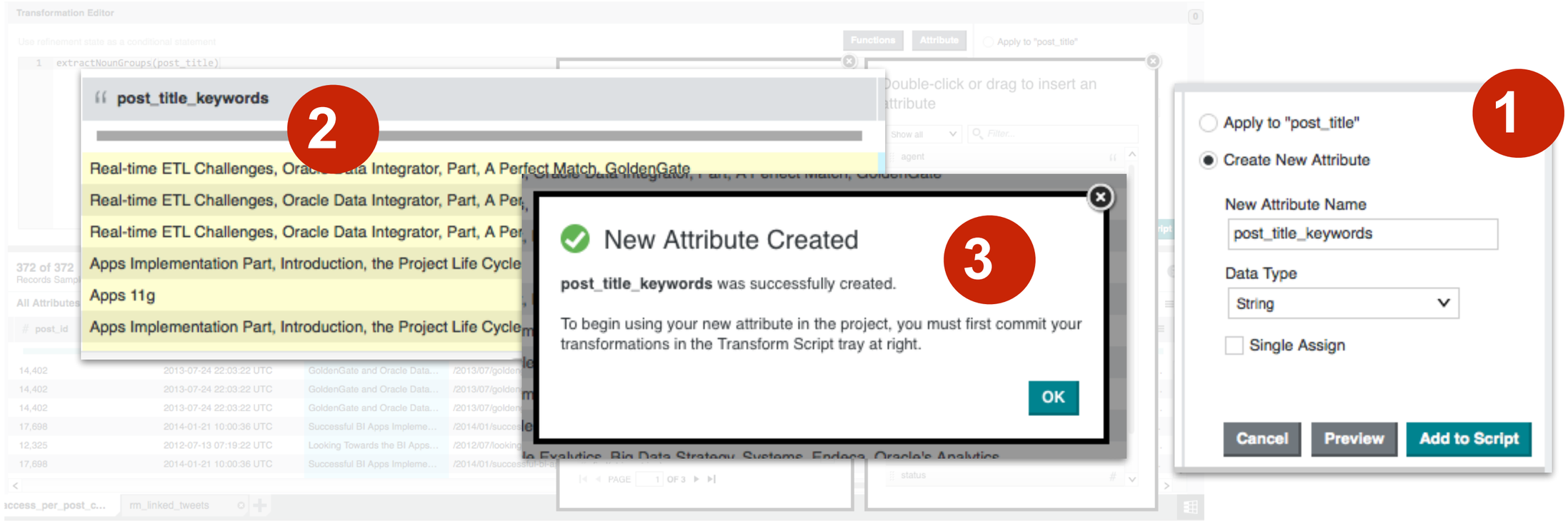
Transformations using Text Enrichment / Parsing

- Uses Salience text engine under the covers
- Extract terms, sentiment, noun groups, positive / negative words etc



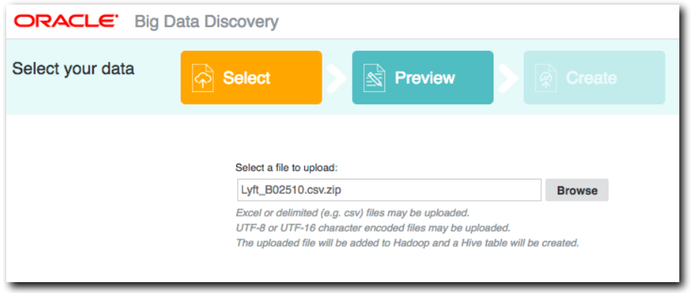
Create New Attribute using Derived (Transformed) Values

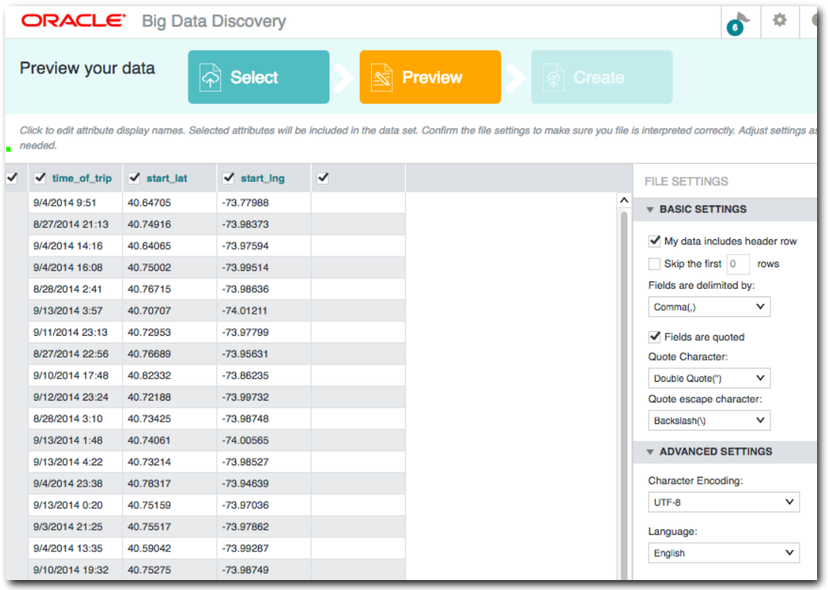
- Choose option to Create New Attribute, to add derived attribute to dataset
- Preview changes, then save to transformation script

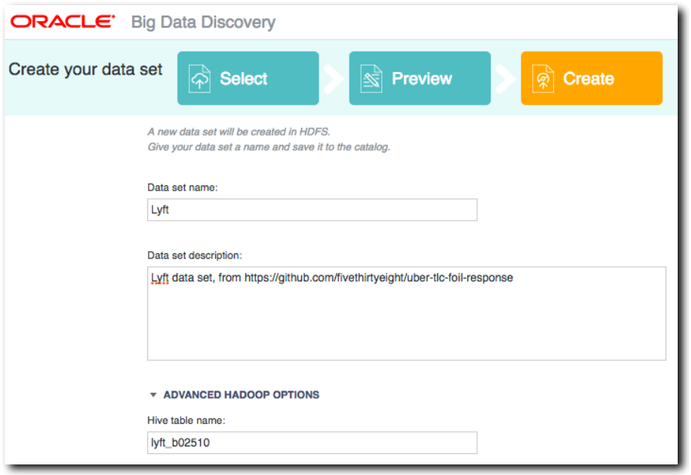


Ingesting Additional Data from File

- Delimited text (such as CSV), or Excel
  - Can be compressed
- Specify delimiter, column names, etc
- Stores the data in HDFS, creates Hive Catalog entry for it, and ingests it to DGraph

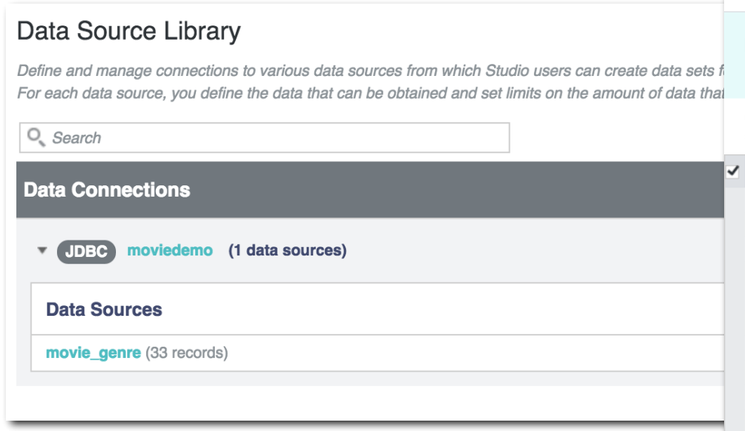


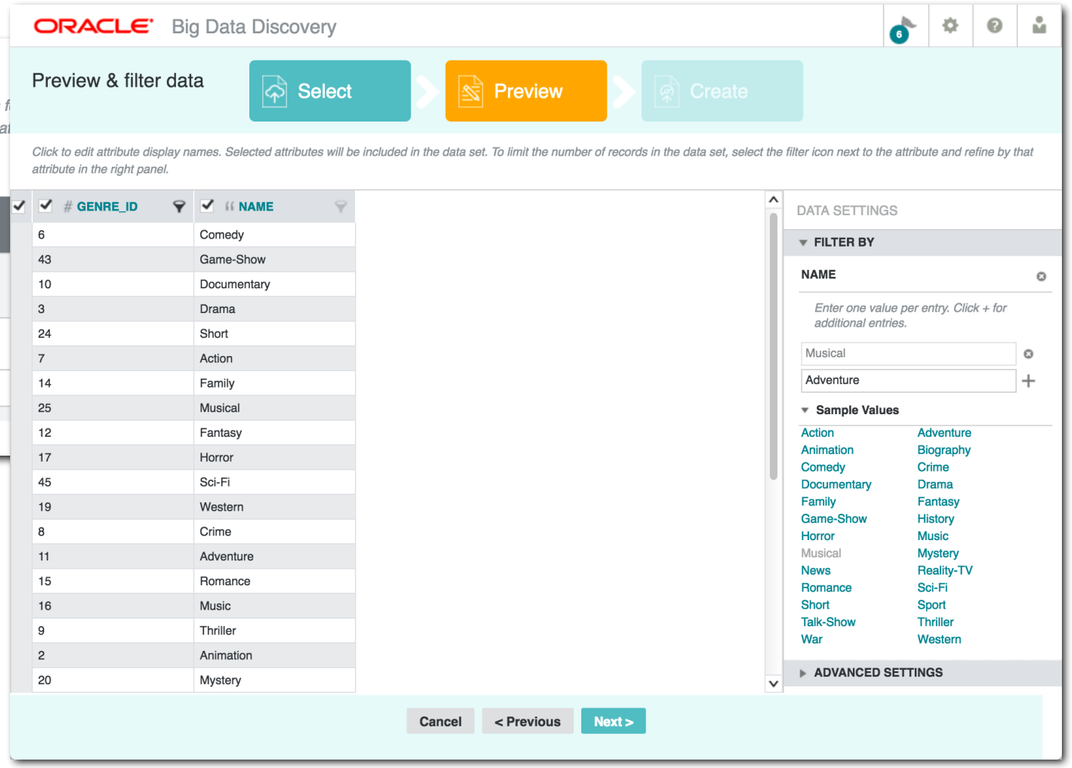


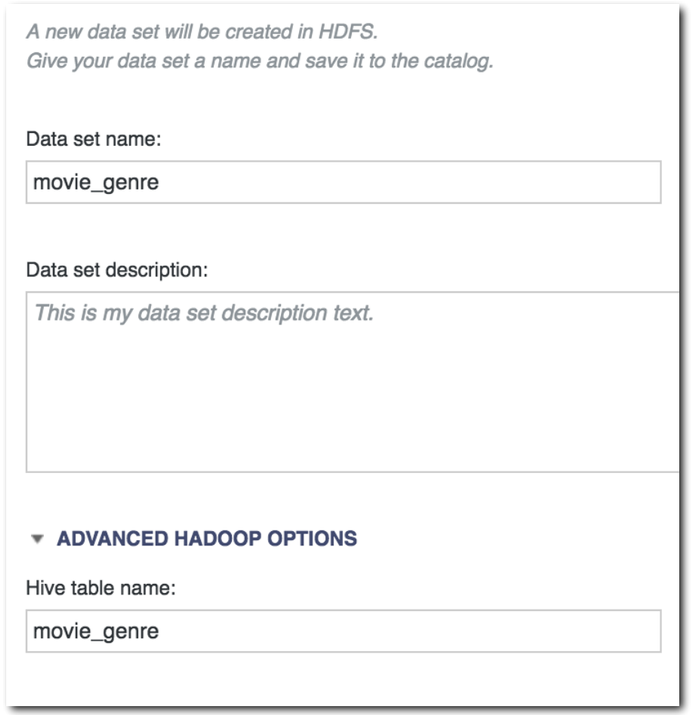


Ingesting Additional Data with JDBC

- Oracle and MySQL currently supported
- Can filter data before ingesting it
- Stores the data in HDFS, creates Hive Catalog entry for it, and ingests it to DGraph

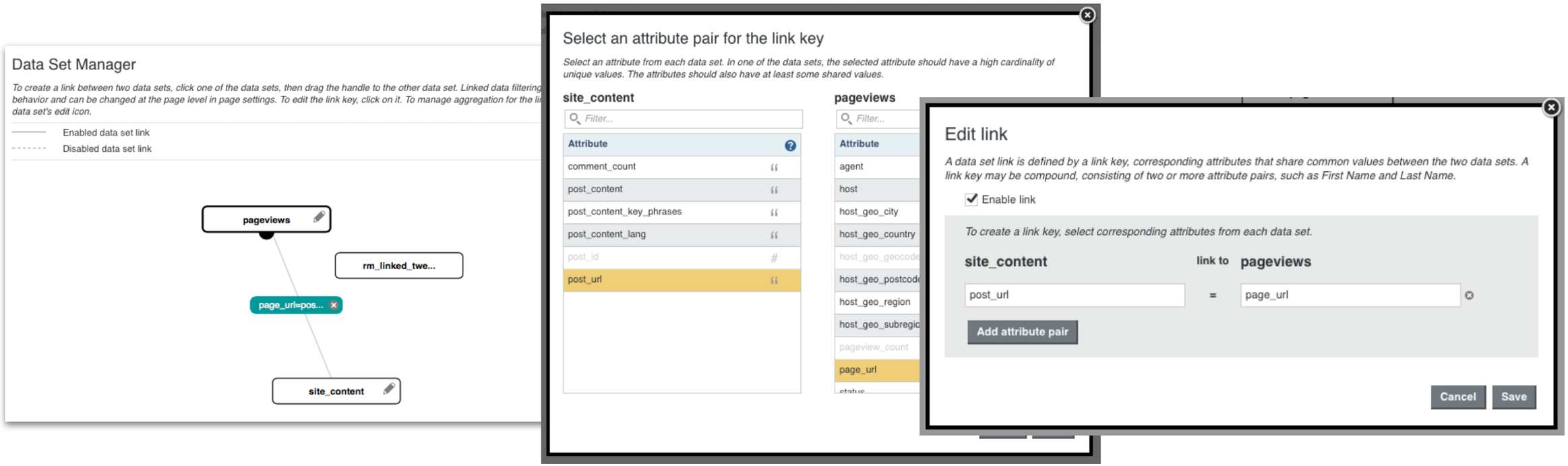






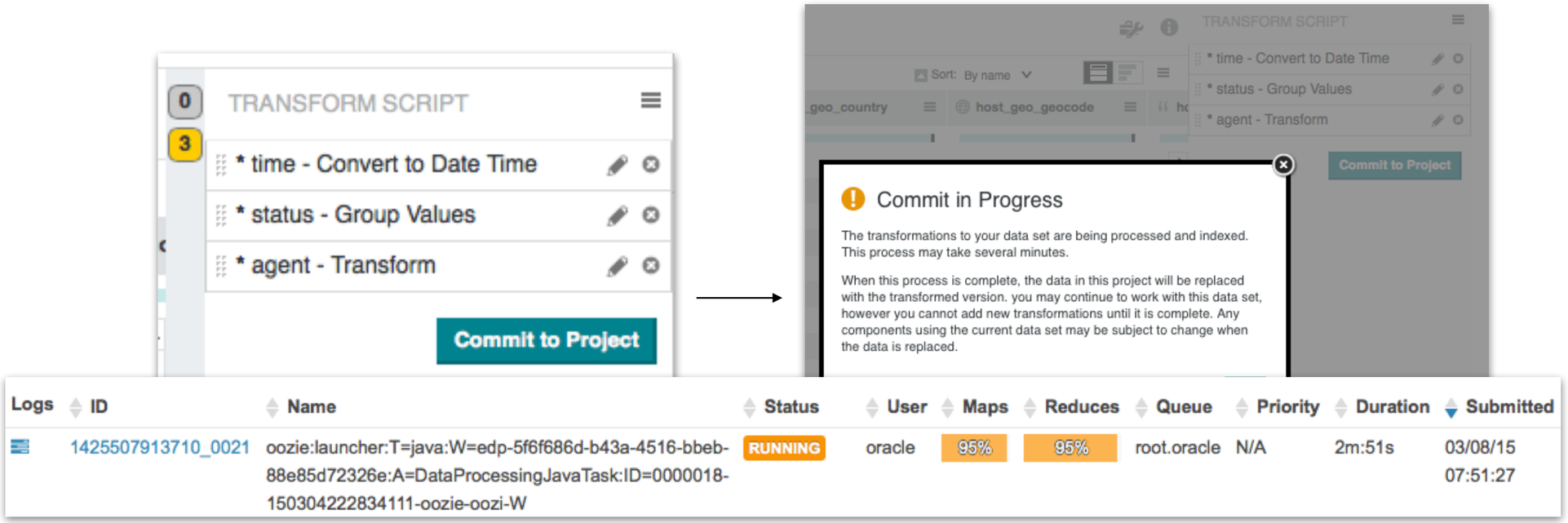
Join Datasets On Common Attributes

- Used to create a dataset based on the intersection (typically) of two datasets
- Not required to just view two or more datasets together - think of this as a JOIN and SELECT



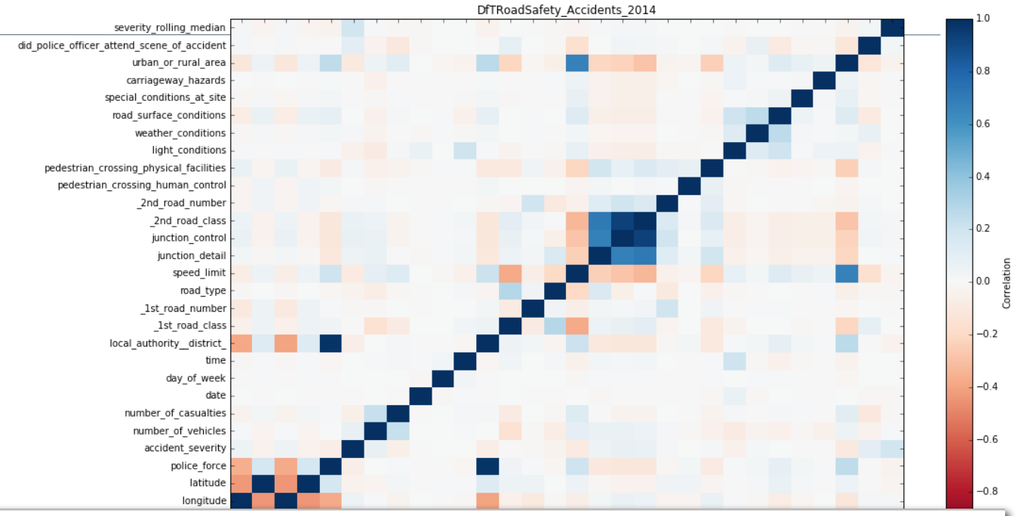
Commit Transforms to DGraph, or Create New Hive Table

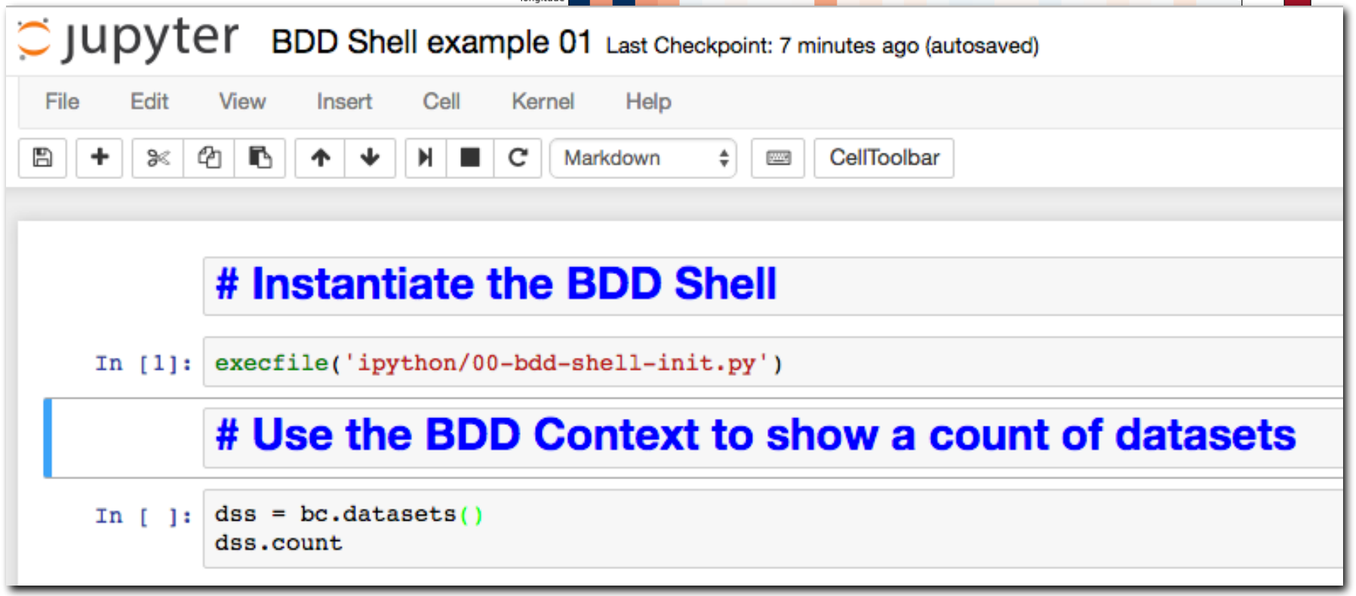
- Transformation changes have to be committed to DGraph sample of dataset
  - Project transformations kept separate from other project copies of dataset
- Transformations can also be applied to full dataset, using Apache Spark
  - Creates new Hive table of complete dataset
- Option to export datasets, locally or to HDFS in Avro or delimted format



BDD Shell and Jupyter Notebooks

- New in BDD 1.2
- Exposes functionality of BDD to Python shell
- Access existing BDD datasets for processing and enrichment in Python/Spark
  - eg Machine Learning, pandas, etc
- Save results of Python/Spark into Hive for subsequent ingest into BDD
- Additional ingest route

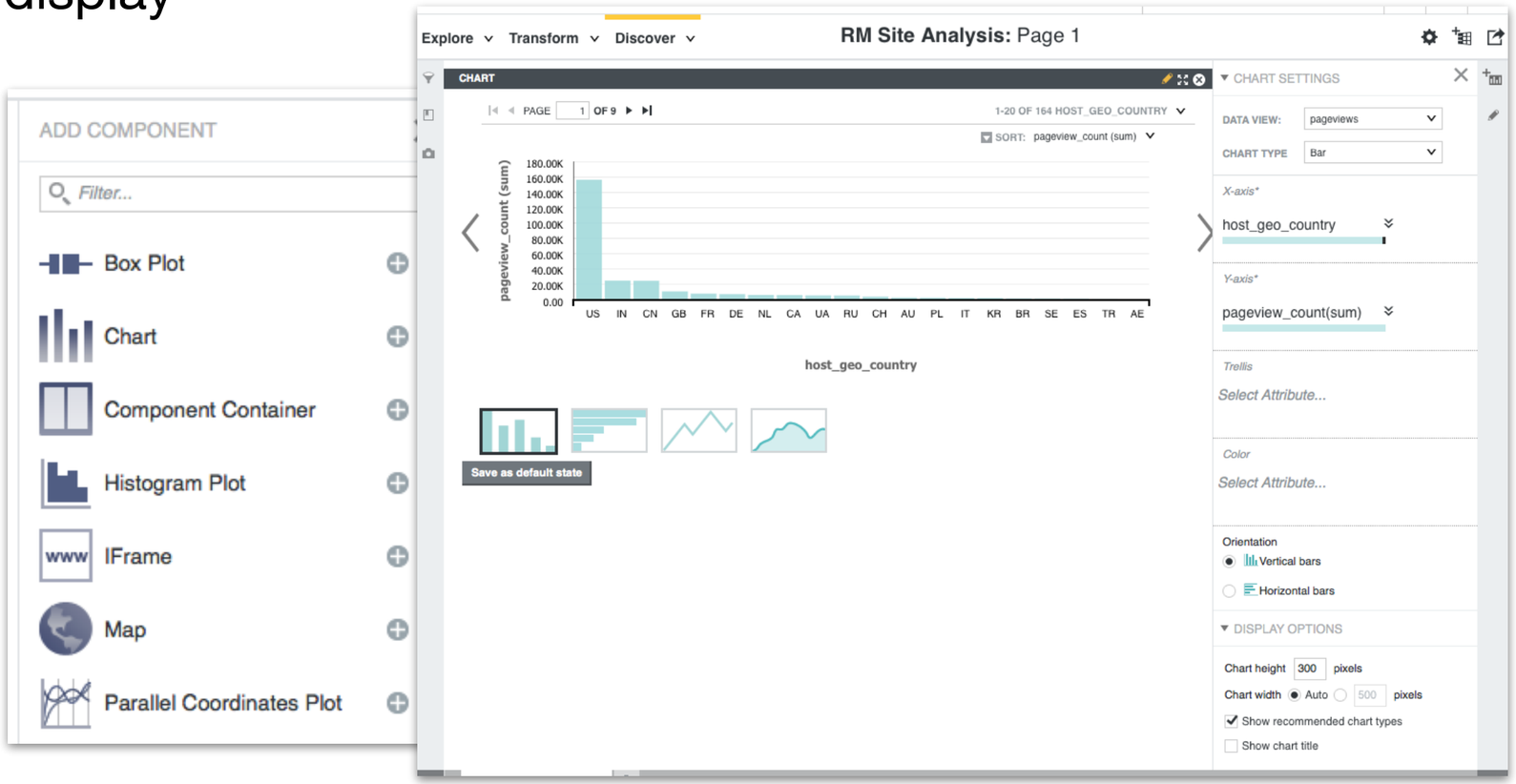




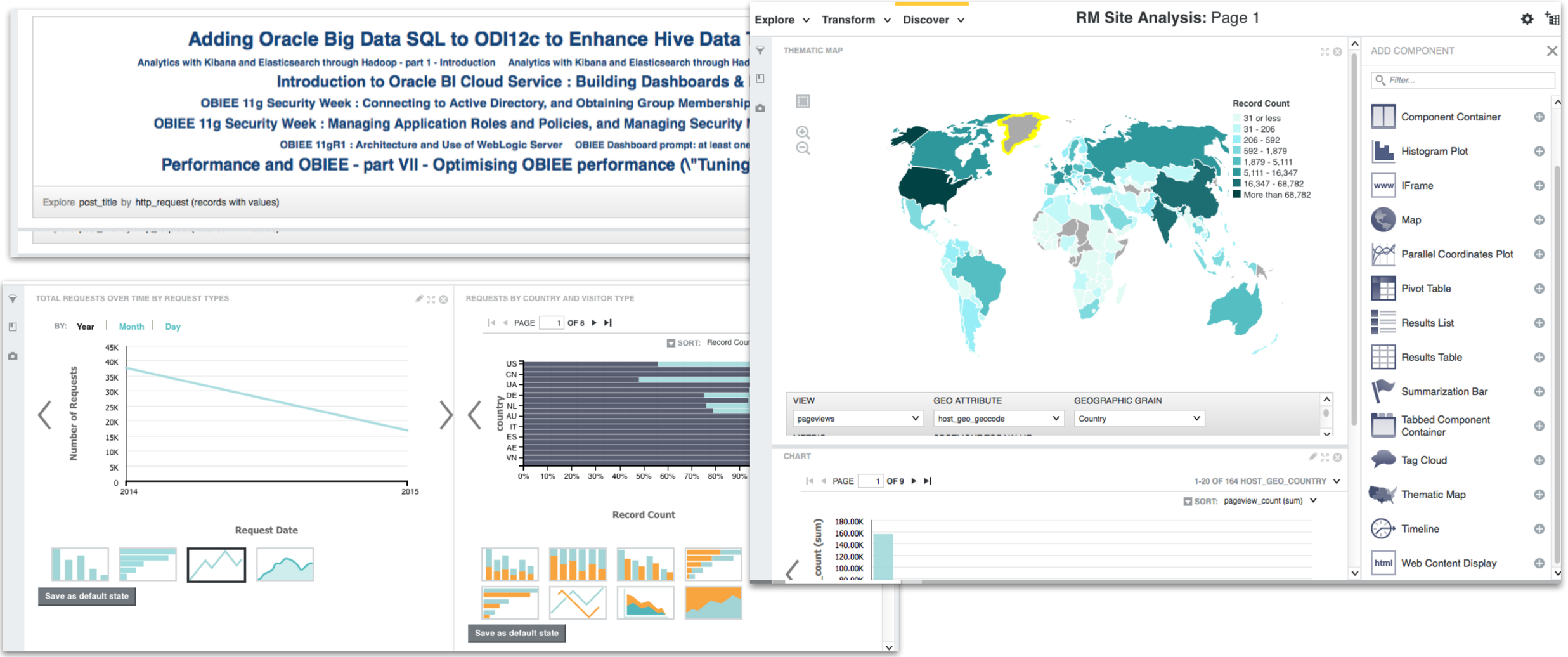
Demo - Big Data Discovery
Data Ingest, Exploration, and Transformation

Create Discovery Pages for Dataset Analysis

- Select from palette of visualisation components
- Select measures, attributes for display

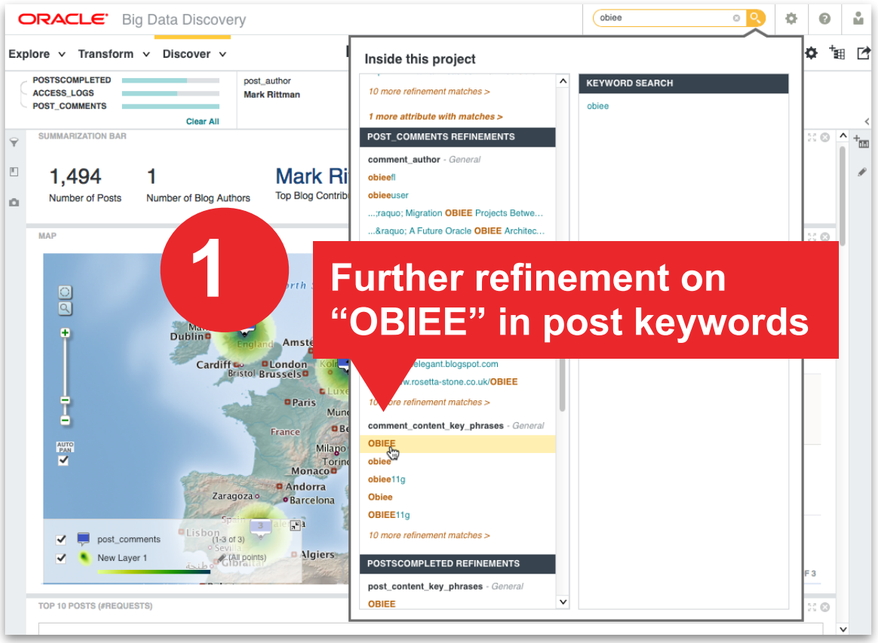


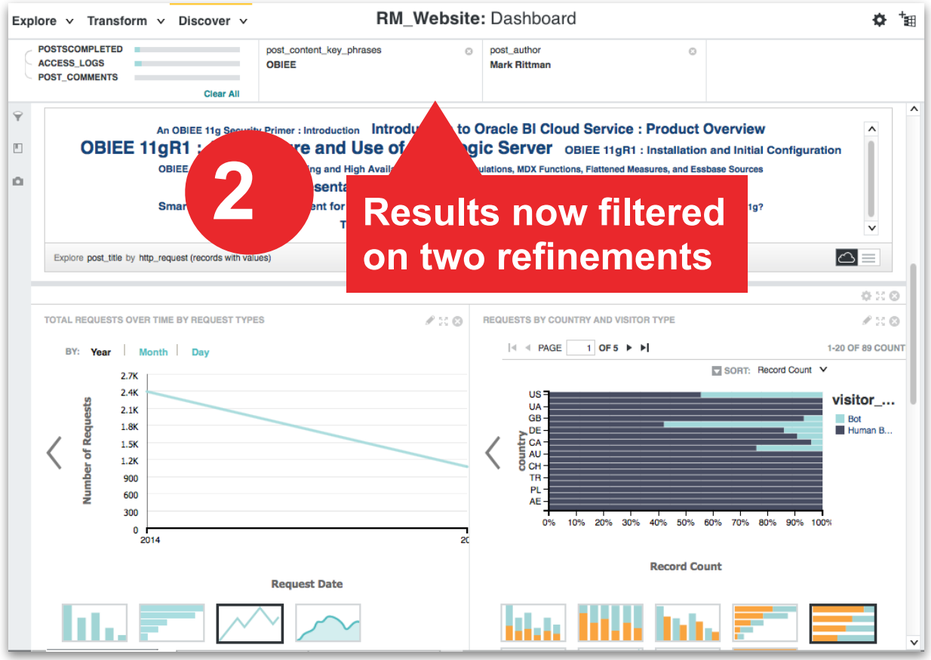
Visualize and Interact With Hadoop Datasets



Faceted Search Across Entire Data Reservoir

- BDD Studio dashboards support faceted search across all attributes, refinements
- Auto-filter dashboard contents on selected attribute values - for data discovery
- Fast analysis and summarisation through Endeca Server technology

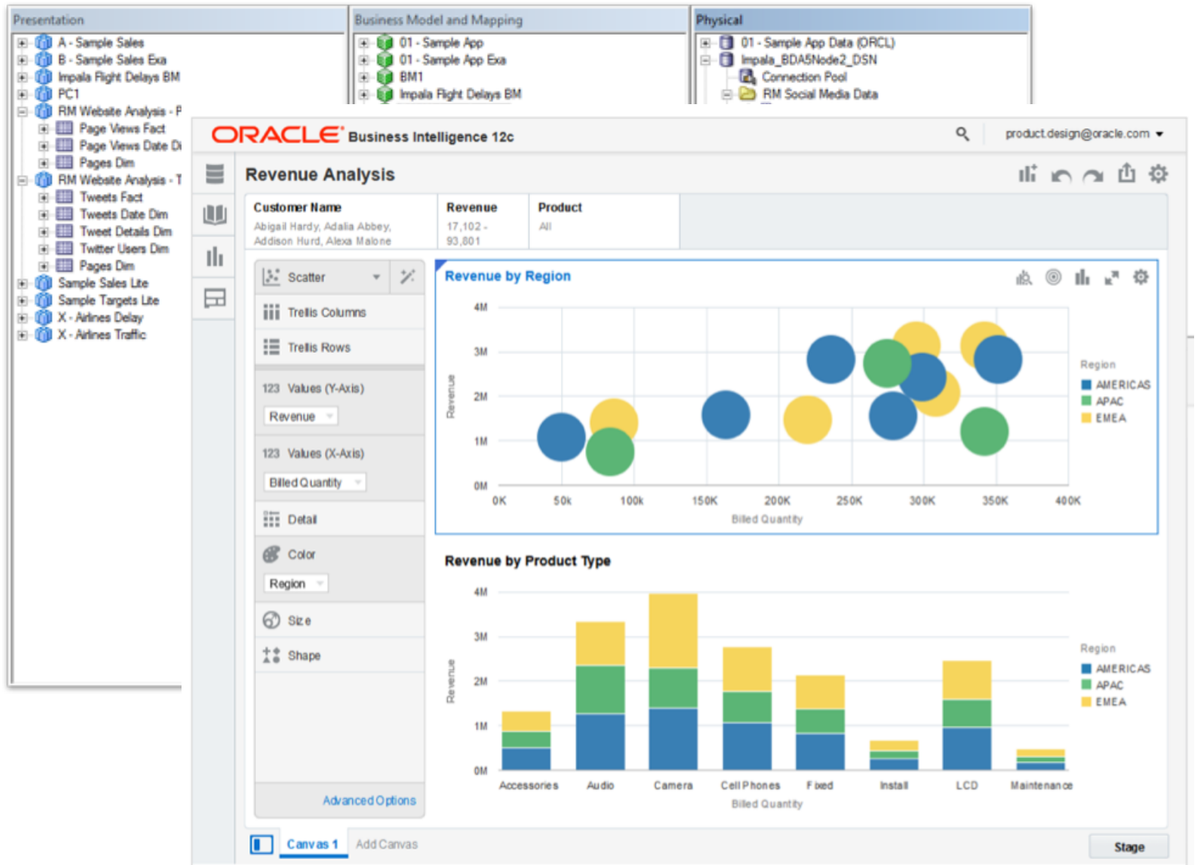




Demo - Big Data Discovery Dashboards

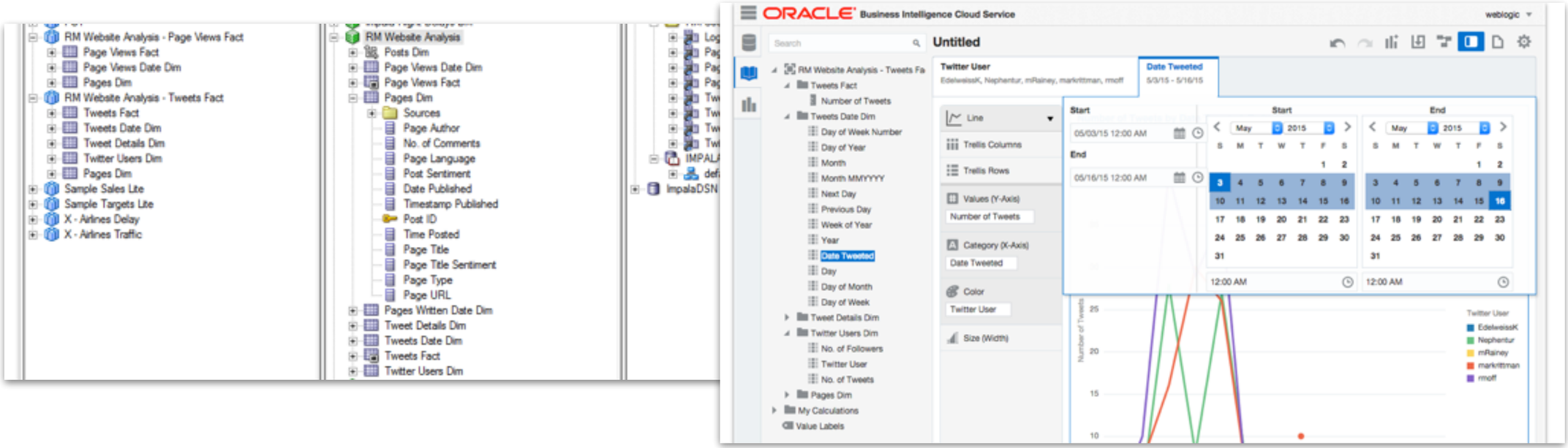
Comparing BDD to Oracle Visual Analyzer

- Visual Analyzer also provides a form of “data discovery” for BI users
  - Similar to Tableau, Qlikview etc
  - Inspired by BI elements of OEID
- Uses OBIEE RPD as the primary datasource, so data needs to be curated + structured
- Probably a better option for users who aren’t concerned it’s “big data”
- But can still connect to Hadoop via Hive, Impala and Oracle Big Data SQL



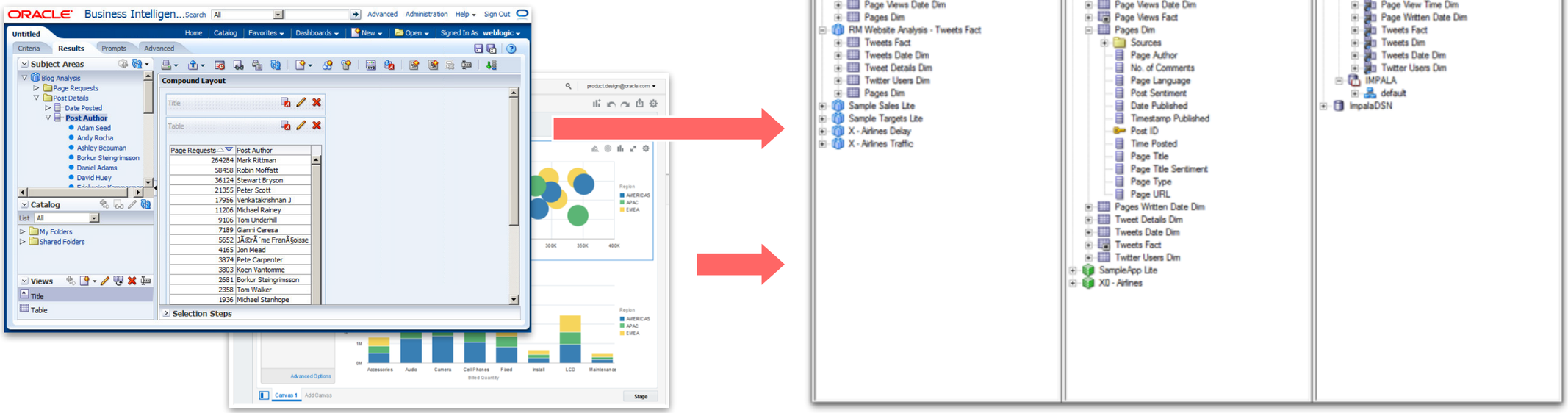
Managed vs. Free-Form Data Discovery

- Data in the data reservoir typically is raw, hasn't been organised into facts, dimensions yet
- In this initial phase, you don't want to it to be - too much up-front work with unknown data
- Later on though, users will benefit from structure and hierarchies being added to data
- But this takes work, and you need to understand cost/benefit of doing it now vs. later



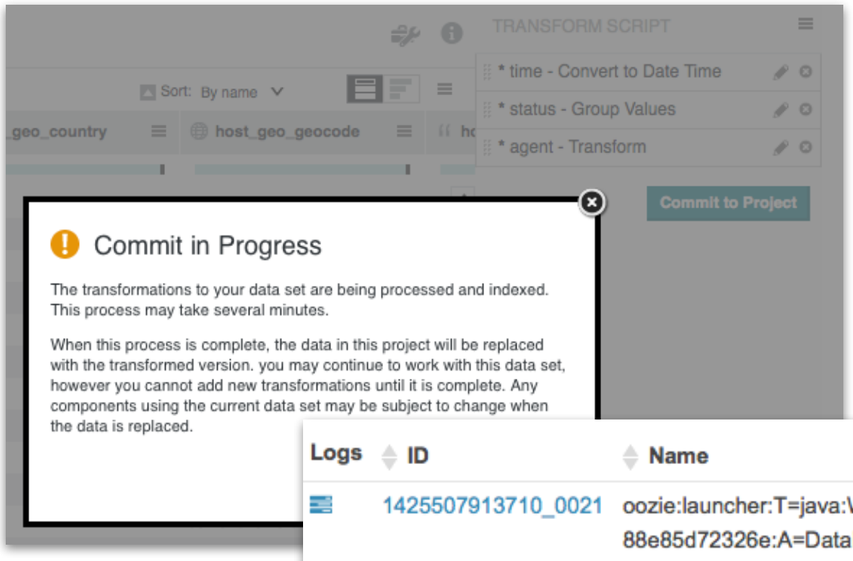
Understand the Work Involved in Creating an RPD

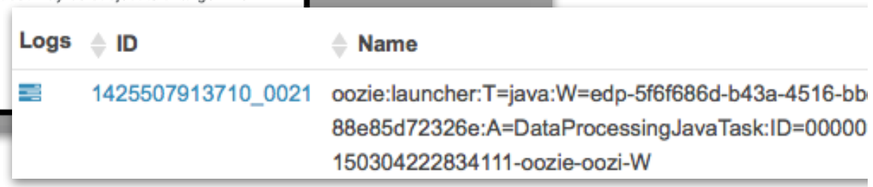
- Visual Analyzer and Answers both require a BI Repository (RPD) as their main datasource
  - Provides a structured, curated baseline for reporting, can be supplemented by mashups
- But is this the right time to be curating data?
  - Do we understand it well enough yet?
  - Do we really need to be modelling it yet?

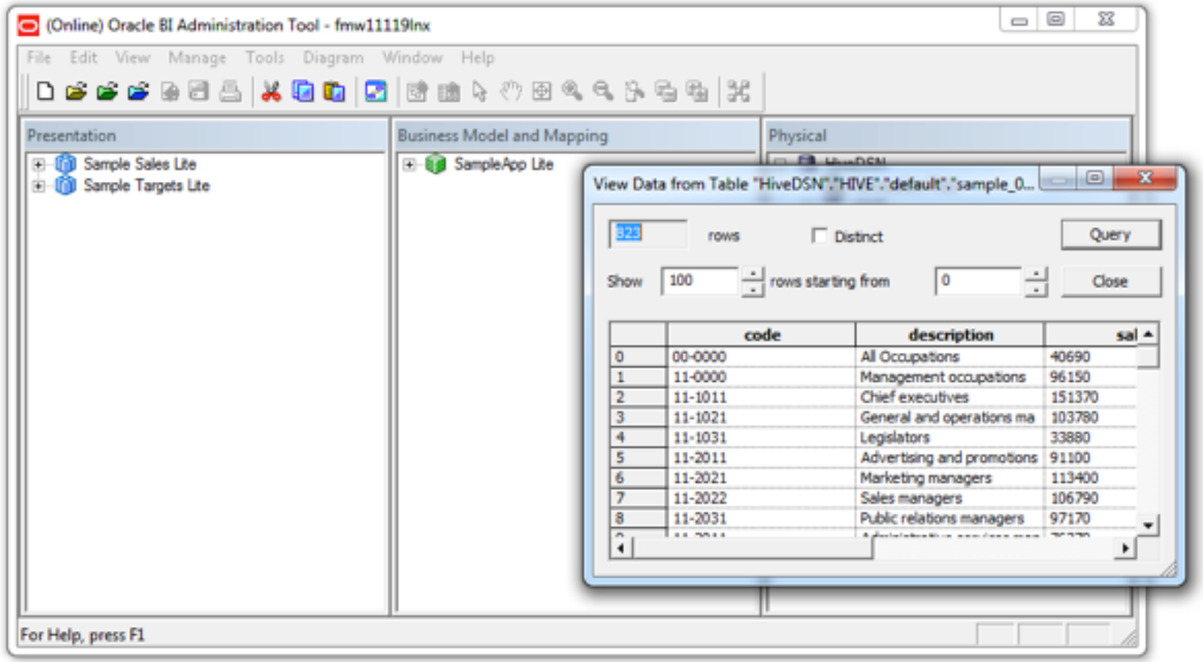


Export Onboard Datasets Back to Hive, for OBIEE + VA

- Transformations within BDD Studio can then be used to create curated fact + dim Hive tables
- Can be used then as a more suitable dataset for use with OBIEE RPD + Visual Analyzer
- Or exported into Exadata or Exalytics to combine with main DW datasets

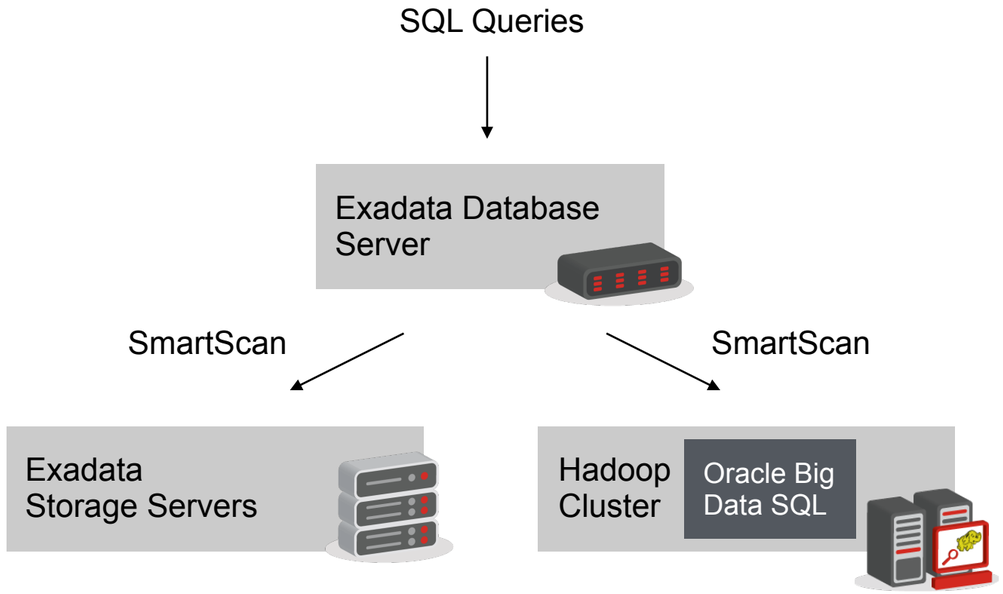






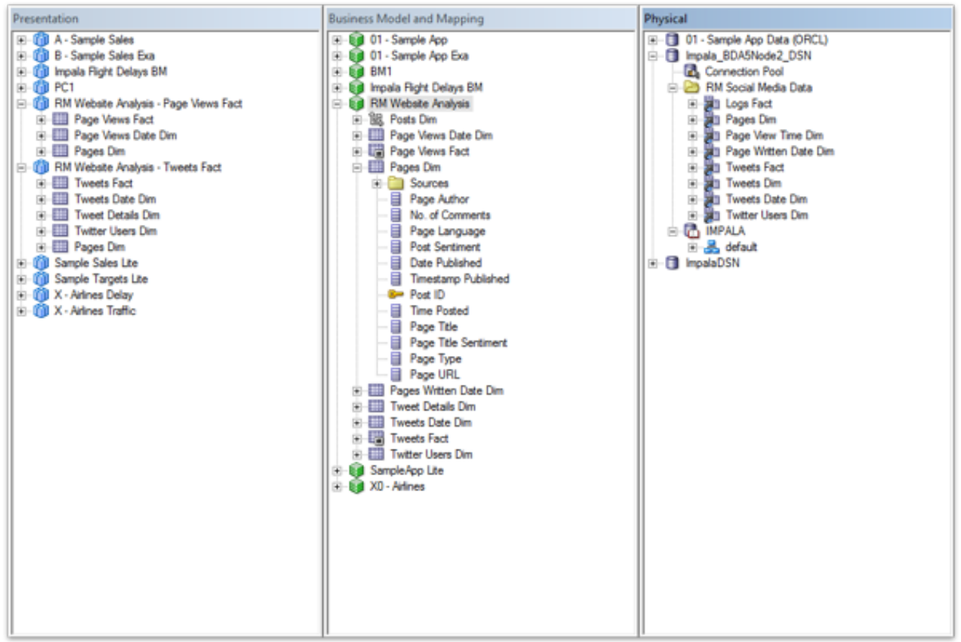
Oracle Big Data SQL

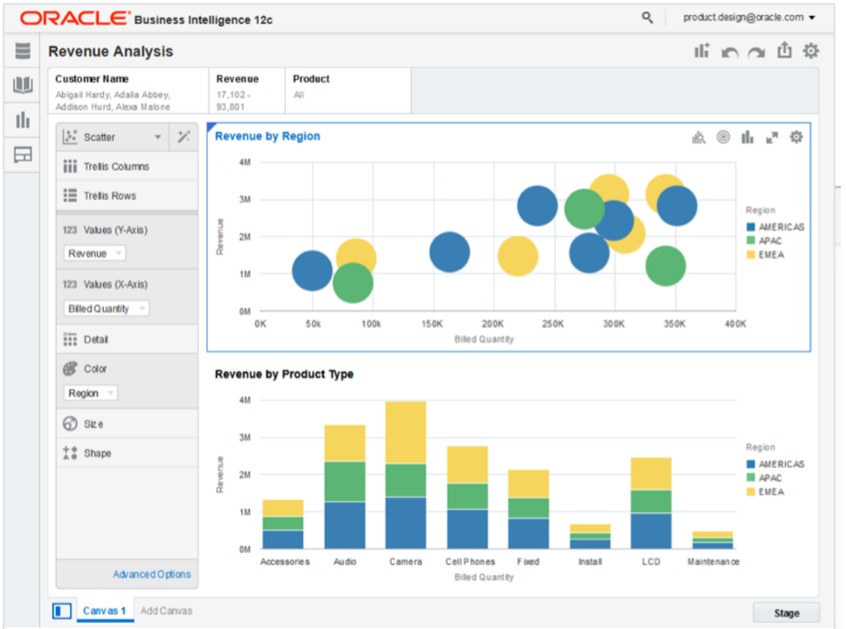
- Part of Oracle Big Data 4.0 (BDA-only)
  - Also requires Oracle Database 12c, Oracle Exadata Database Machine
- Extends Oracle Data Dictionary to cover Hive
- Extends Oracle SQL and SmartScan to Hadoop
- Extends Oracle Security Model over Hadoop
  - Fine-grained access control
  - Data redaction, data masking
- Uses fast c-based readers where possible (vs. Hive MapReduce generation)
- Map Hadoop parallelism to Oracle PQ
- Big Data SQL engine works on top of YARN
  - Like Spark, Tez, MR2



Create the RPD Against Curated, Enriched Hive Tables

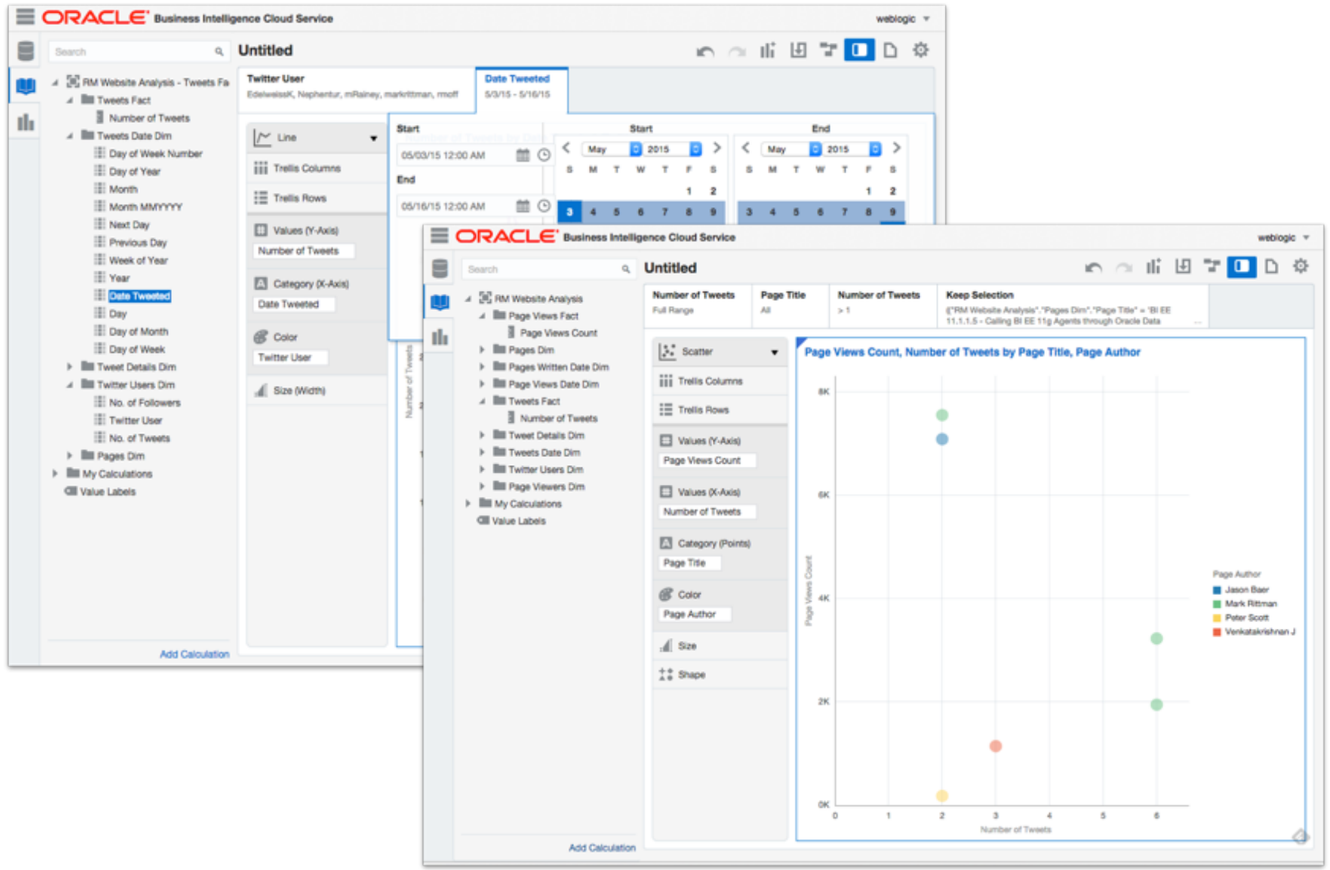
- Now is the time to invest time into creating the RPD
- We understand the data, have added enrichments, discovered the hierarchies
- The next set of users will benefit from time taken to curate the data into an RPD

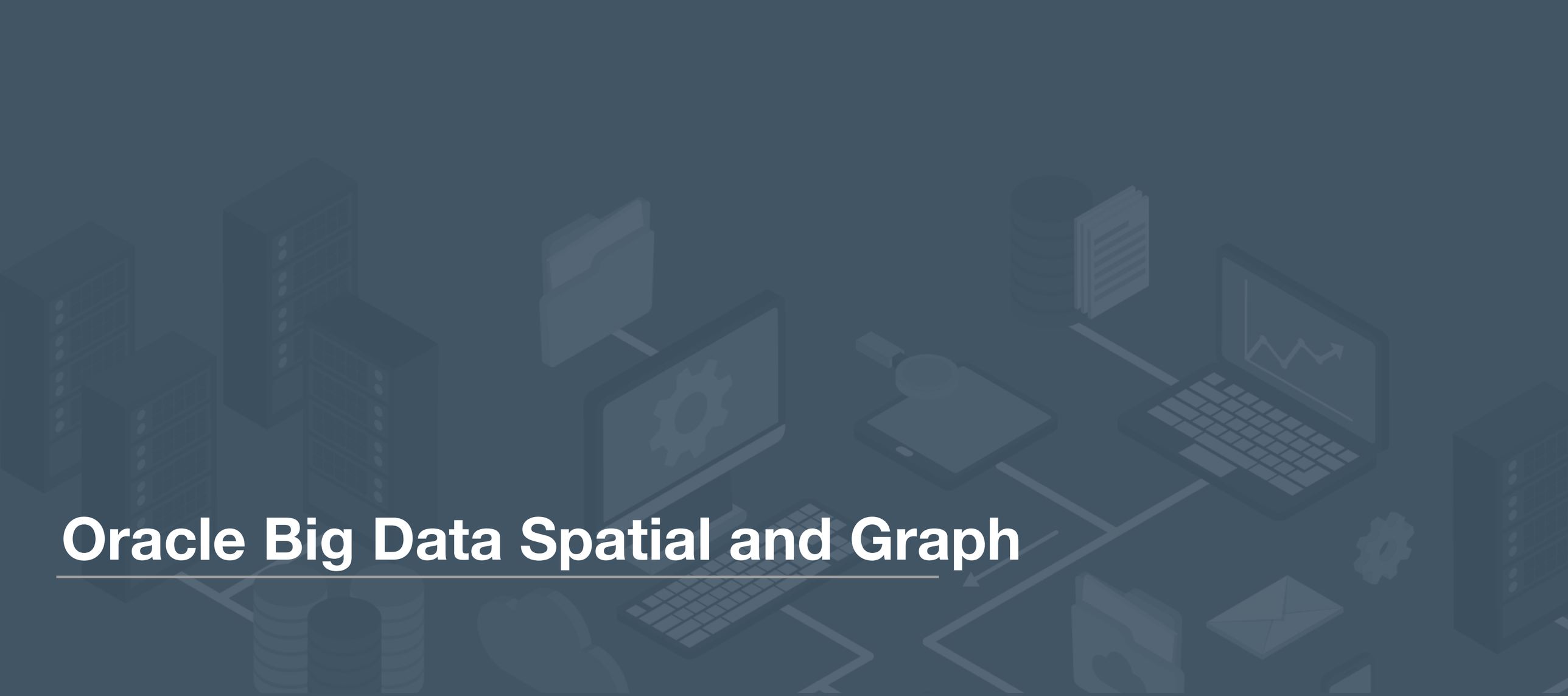




Further Analyse in Visual Analyzer for Managed Dataset

- Users in Visual Analyzer then have a more structured dataset to use
- Data organised into dimensions, facts, hierarchies and attributes
- Can still access Hadoop directly through Impala or Big Data SQL
- Big Data Discovery though was key to initial understanding of data





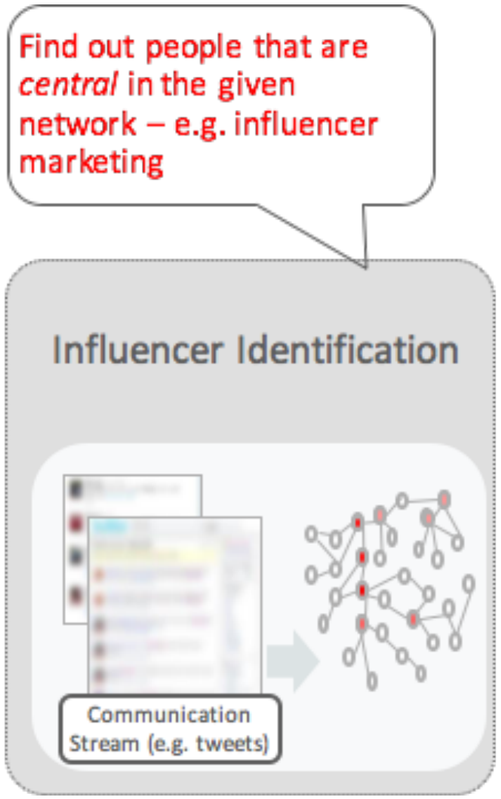
Oracle Big Data Spatial and Graph

info@rittmanmead.com www.rittmanmead.com @rittmanmead

rittmanmead

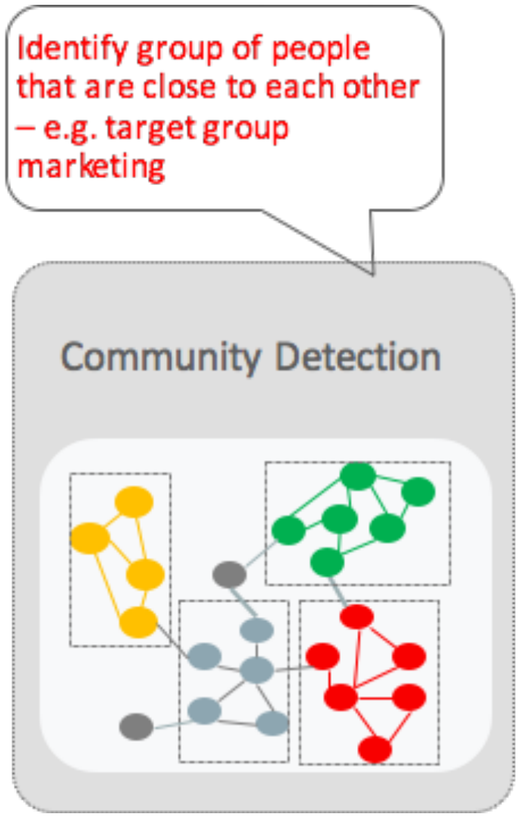
Who Are The Influencers In Our Community?

- Sometimes the highest number isn't the most important
- For example, some Twitter users are far more influential than others
  - Sit at the centre of a community, have 1000's of followers
  - A reference by them has massive impact on page views
  - Positive or negative comments from them drive perception
- Can we identify them?
  - Potentially "reach out" with analyst program
  - Study what website posts go "viral"
  - Understand out audience, and the conversation, better



What Communities and Networks Are Our Audience?

- Rittman Mead website features many types of content
  - Blogs on BI, data integration, big data, data warehousing
  - Op-Eds ("OBIEE12c - Three Months In, What's the Verdict?")
  - Articles on a theme, e.g. performance tuning
  - Details of new courses, new promotions
- Different communities likely to form around these content types
- Different influencers and patterns of recommendation, discovery
- Can we identify some of the communities, segment our audience?



Graph Example : RM Blog Post Referenced on Twitter

Follows

Follows

rittmanmead

Lifting the Lid on OBIEE Internals with Linux Diagnostics Tools

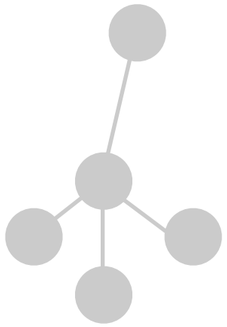
Edit | March 27, 2015 | Oracle BI Suite EE | 2 Comments

There comes the point in any sufficiently complex or difficult problem diagnosis that the log files in OBIEE alone are not sufficient for building up a complete picture of what's going on. Even with the debug/trace data that Presentation Services and other components can be configured precisely to write you're sometimes just left having to guess what is going on inside the black box of each of the OBIEE system components.

Here we're going to look at a couple of examples of lifting the lid just a little bit further on what OBIEE is up to, using standard Linux diagnostic tools. These are not something to be reaching for in the first instance, but more getting on to a last resort. Almost always

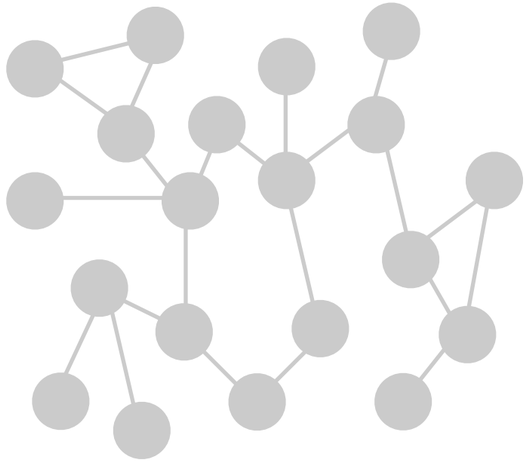
0 0 0 3 Page Views

Network Effect Magnified by Extent of Social Graph



Neselovskyi, Borys TWEETS 552 FOLLOWING 158 FOLLOWERS 76
@bneselov

Lifting the Lid on OBIEE Internals with Linux Diagnostics Tools http://t.co/gFcUPOm5pl



Gwen (Chen) Shapira TWEETS 14.9K FOLLOWING 2,012 FOLLOWERS 5,444 LIKES 1,797
@gwenshap FOLLOWS YOU

Lifting the Lid on OBIEE Internals with Linux Diagnostics Tools http://t.co/gFcUPOm5pl

rittmanmead

Lifting the Lid on OBIEE Internals with Linux Diagnostics Tools

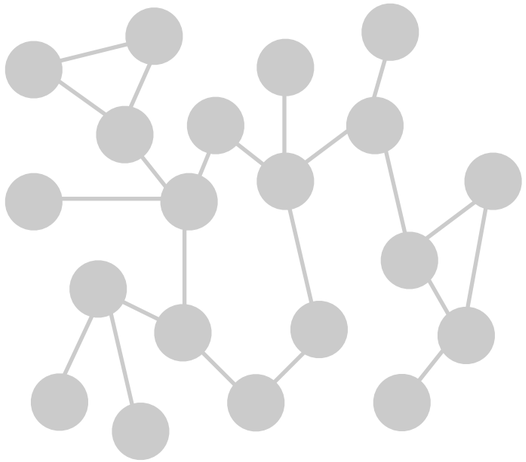
Edit | March 27, 2015 | Oracle BI Suite EE | 2 Comments

There comes the point in any sufficiently complex or difficult problem diagnosis that the log files in OBIEE alone are not sufficient for building up a complete picture of what's going on. Even with the debug/trace data that Presentation Services and other components can be configured precisely to write you're sometimes just left having to guess what is going on inside the black box of each of the OBIEE system components.

Here we're going to look at a couple of examples of lifting the lid just a little bit further on what OBIEE is up to, using standard Linux diagnostic tools. These are not something to be reaching for in the first instance, but more getting on to a last resort. Almost always

0057 Page Views

Retweets by Influential Twitter Users Drive Visits



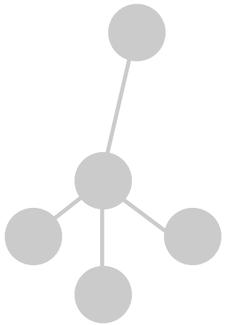
Gwen (Chen) Shapira @gwenshap FOLLOWS YOU

TWEETS	FOLLOWING	FOLLOWERS	LIKES
14.9K	2,012	5,444	1,797

RT: Lifting the Lid on OBIEE Internals with Linux Diagnostics Tools http://t.co/gFcUPOm5pI

0 0 3 5 Page Views

Retweet



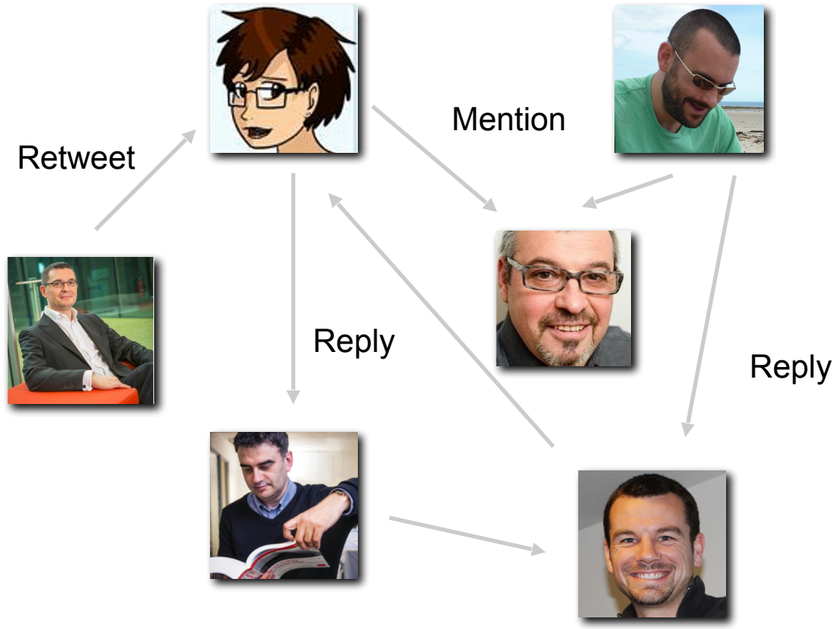
Neselovskyi, Borys @bneselov

TWEETS	FOLLOWING	FOLLOWERS
552	158	76

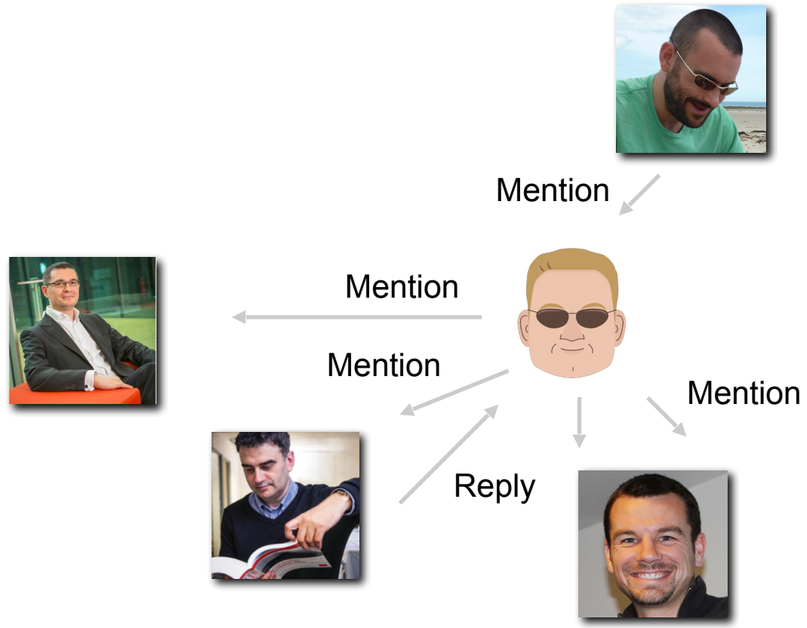
Lifting the Lid on OBIEE Internals with Linux Diagnostics Tools http://t.co/gFcUPOm5pI

0 0 0 3 Page Views

Retweets, Mentions and Replies Create Communities

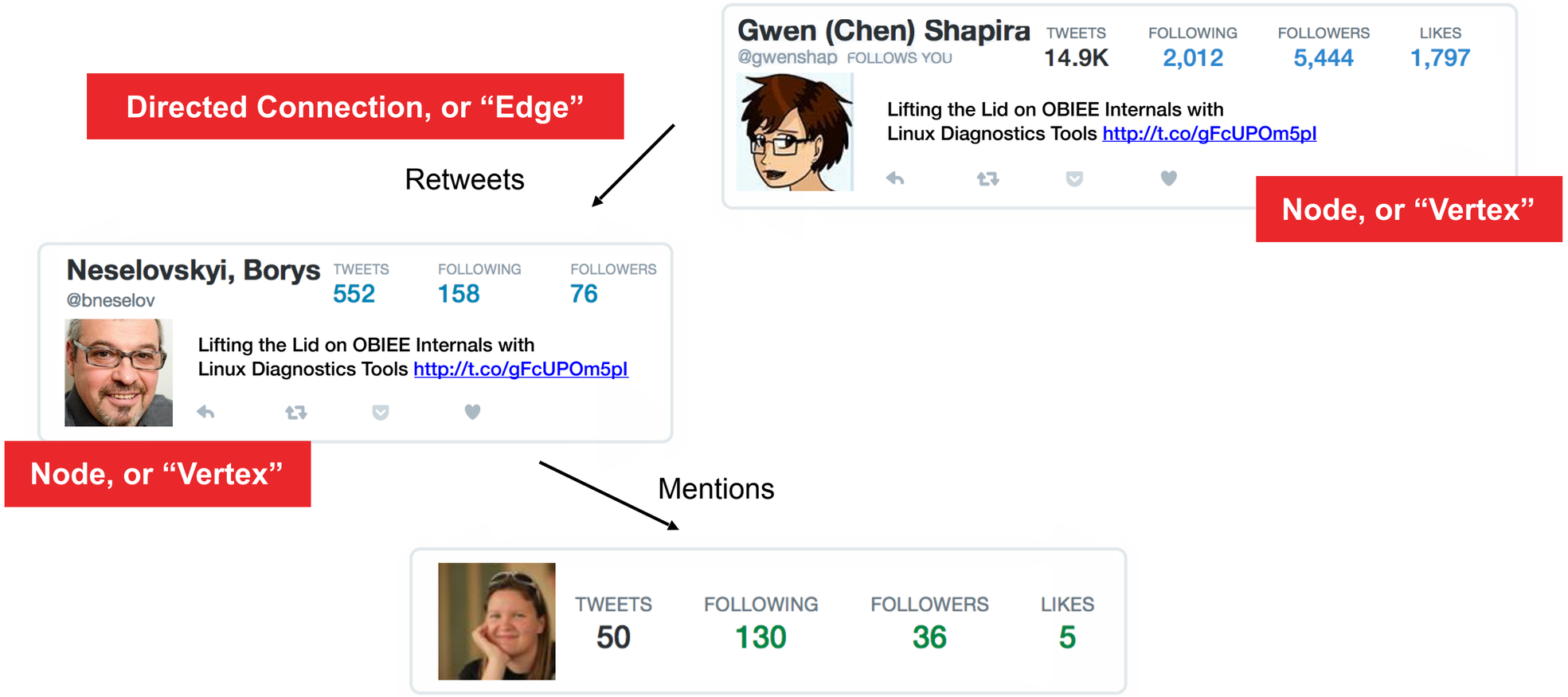


#bigdatasql



#thatswhatshesaid

Property Graph Terminology



Determining Influencers - Factors to Consider

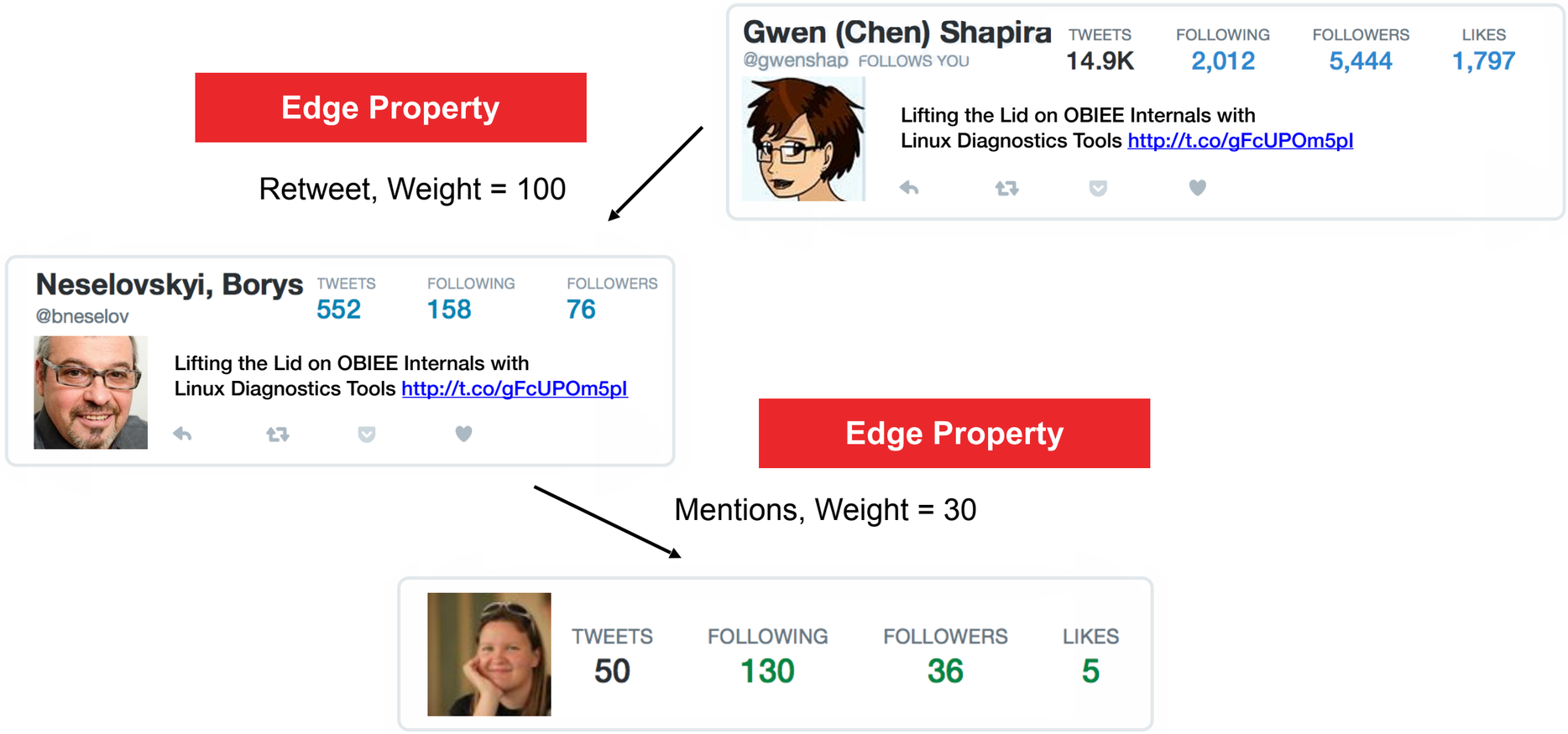
- Different types of Twitter interaction could imply more or less “influence”

  - **Retweet** of another user’s Tweet implies that person is worth quoting or you endorse their opinion

  - **Reply to** another user’s tweet could be a weaker recognition of that person’s opinion or view

  - **Mention** of a user in a tweet is a weaker recognition that they are part of a community / debate

Relative Importance of Edge Types Added via Weights



Oracle Big Data Spatial & Graph

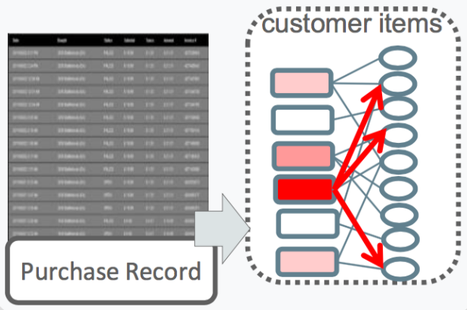
- Graph, spatial and raster data processing for big data
  - Runs on-prem, or in Oracle Big Data Cloud Service
  - Installable on commodity cluster using CDH
- Data stored in Apache HBase or Oracle NoSQL DB
  - Complements Spatial & Graph in Oracle Database
  - Designed for trillions of nodes, edges etc
- Out-of-the-box spatial enrichment services
- Over 35 of most popular graph analysis functions
  - Graph traversal, recommendations
  - Finding communities and influencers,
  - Pattern matching

Graph Analysis Uses

Common Graph Analysis Use Cases

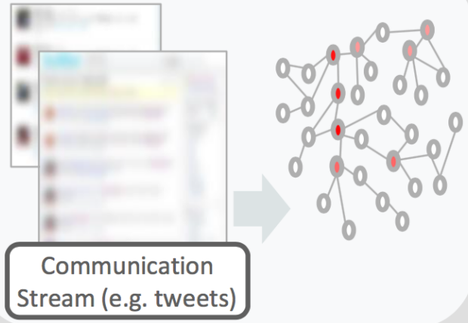
Recommend the most *similar* item purchased by *similar* people

Product Recommendation



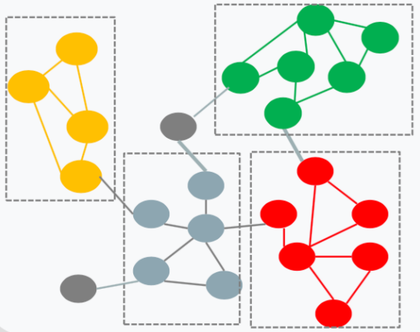
Find out people that are *central* in the given network – e.g. influencer marketing

Influencer Identification



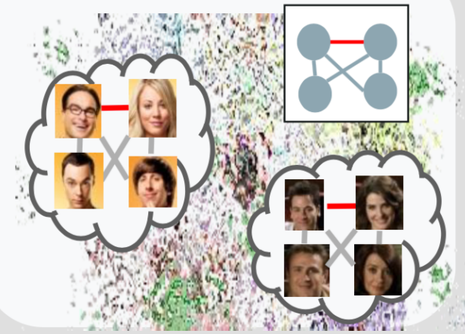
Identify group of people that are close to each other – e.g. target group marketing

Community Detection

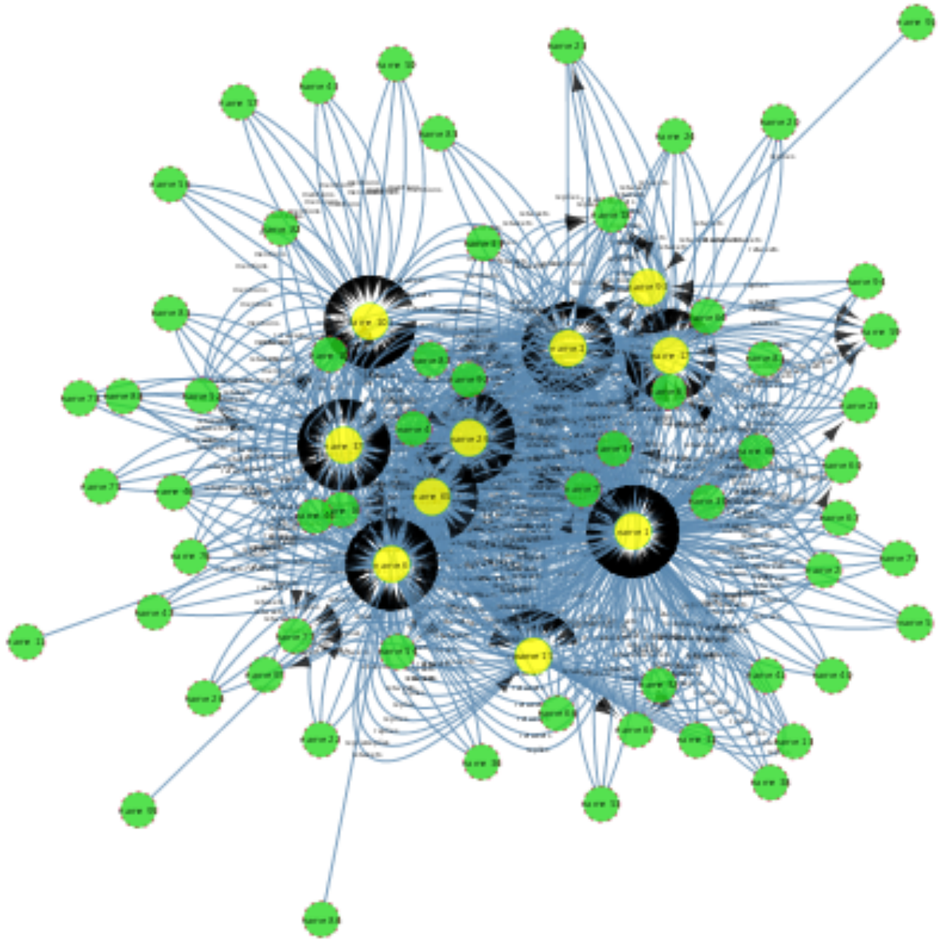


Find out all the sets of entities that match to the given pattern – e.g. fraud detection

Graph Pattern Matching



Calculating Top 10 Users using Page Rank Algorithm



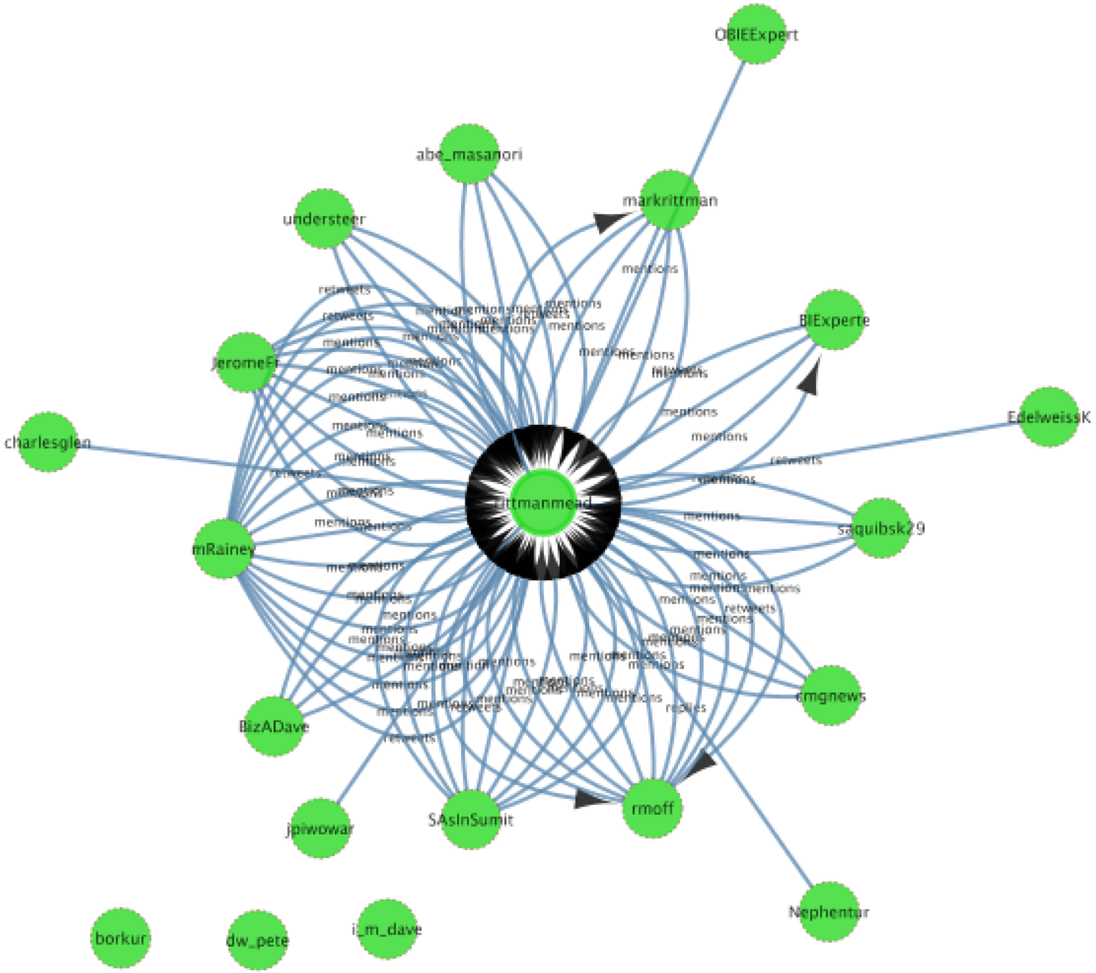
In-memory Analytics Starting Point:

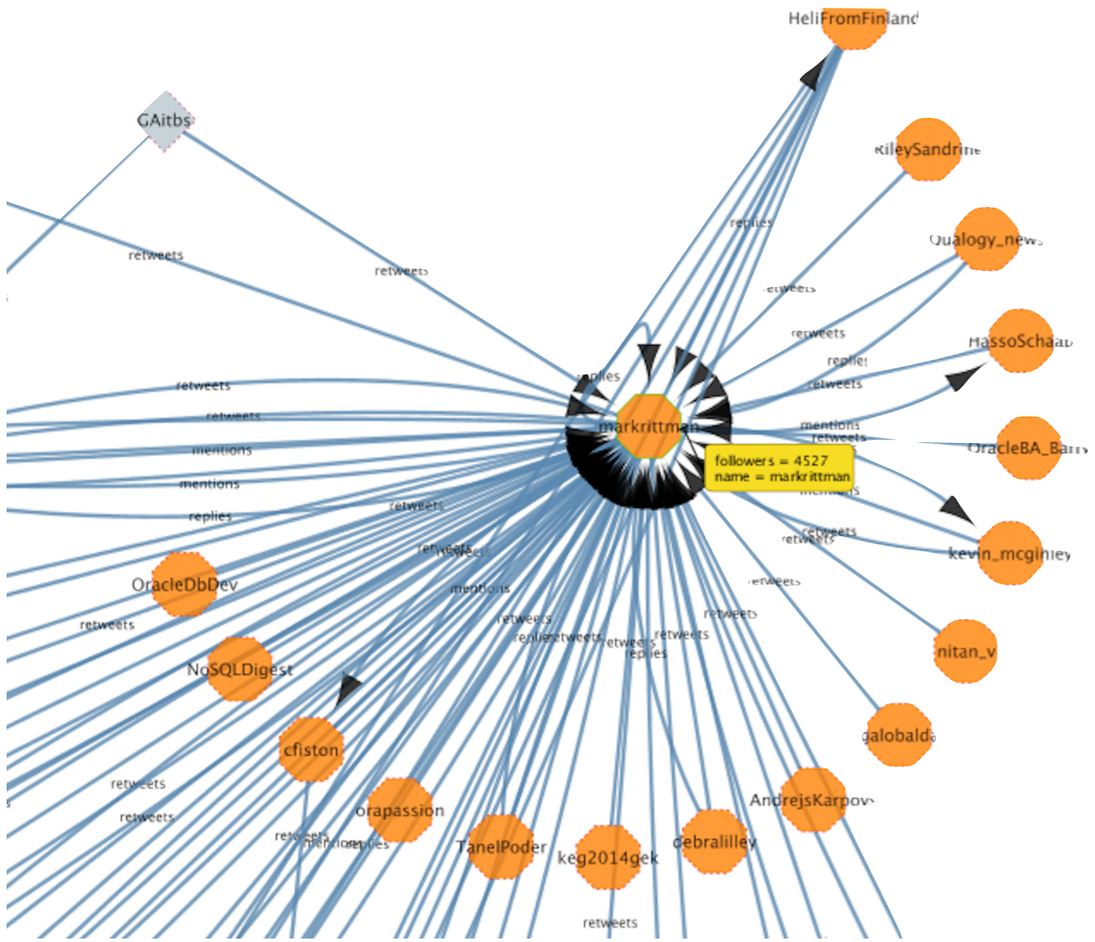
- (•) page rank
- ( ) in degree centrality
- ( ) out degree centrality
- ( ) vertex betweenness centrality

Top-ranked: 10

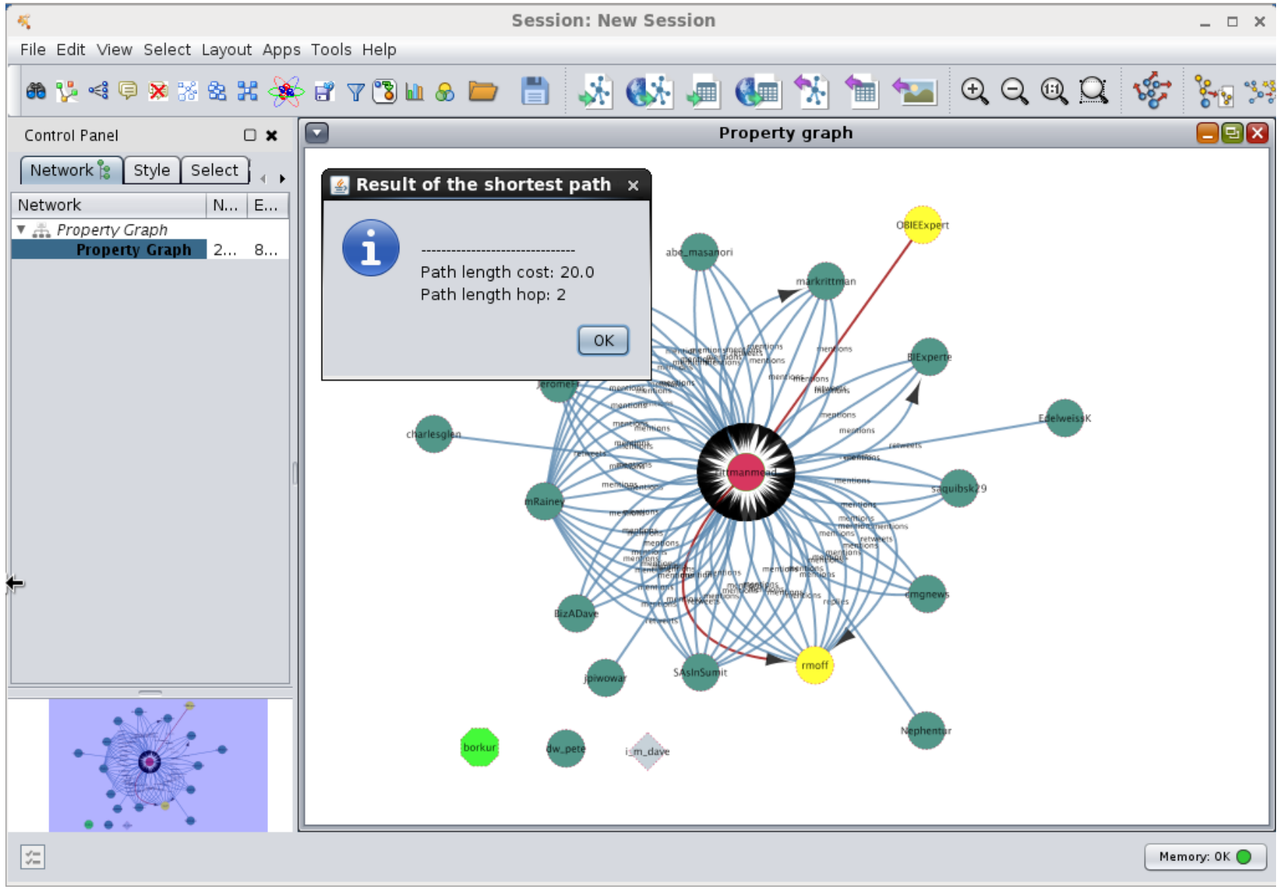
```
Top 10 influencers:
 markrittman
 rmoff
 rittmanmead
 mRainey
 JeromeFr
 Nephentur
 borkur
 BIExperte
 i_m_dave
 dw_pete
```

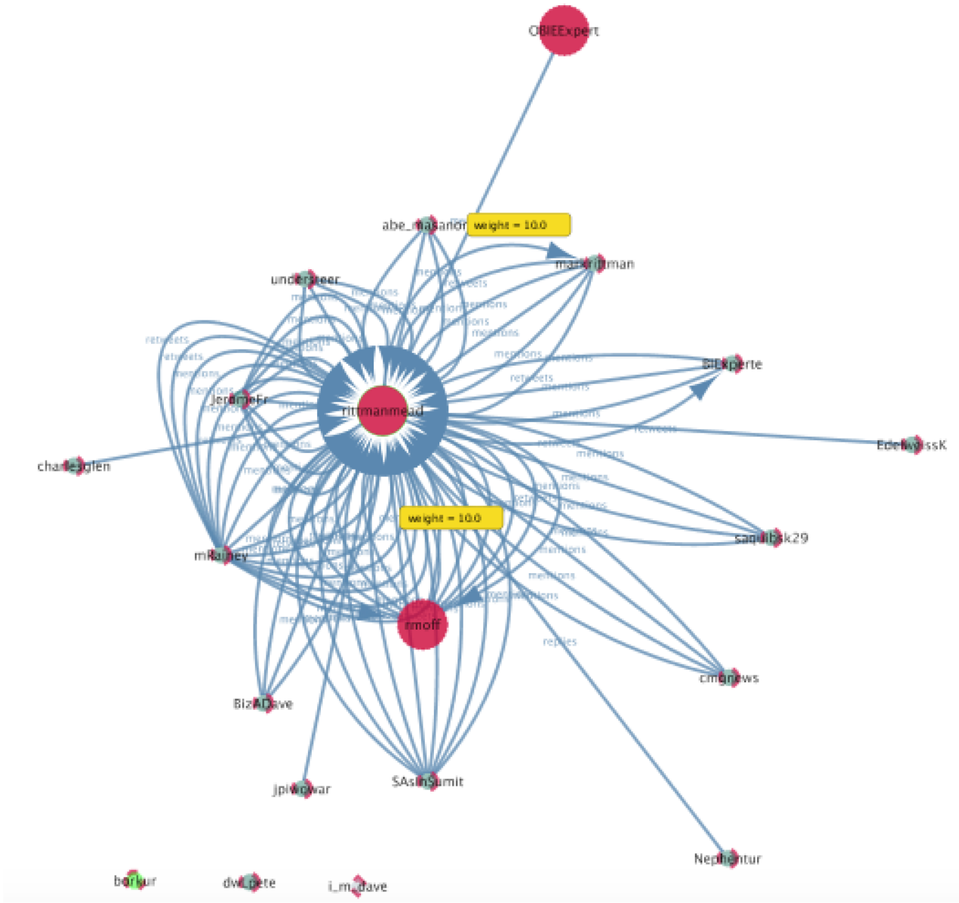
Visualising the Social Graph Around Particular Users



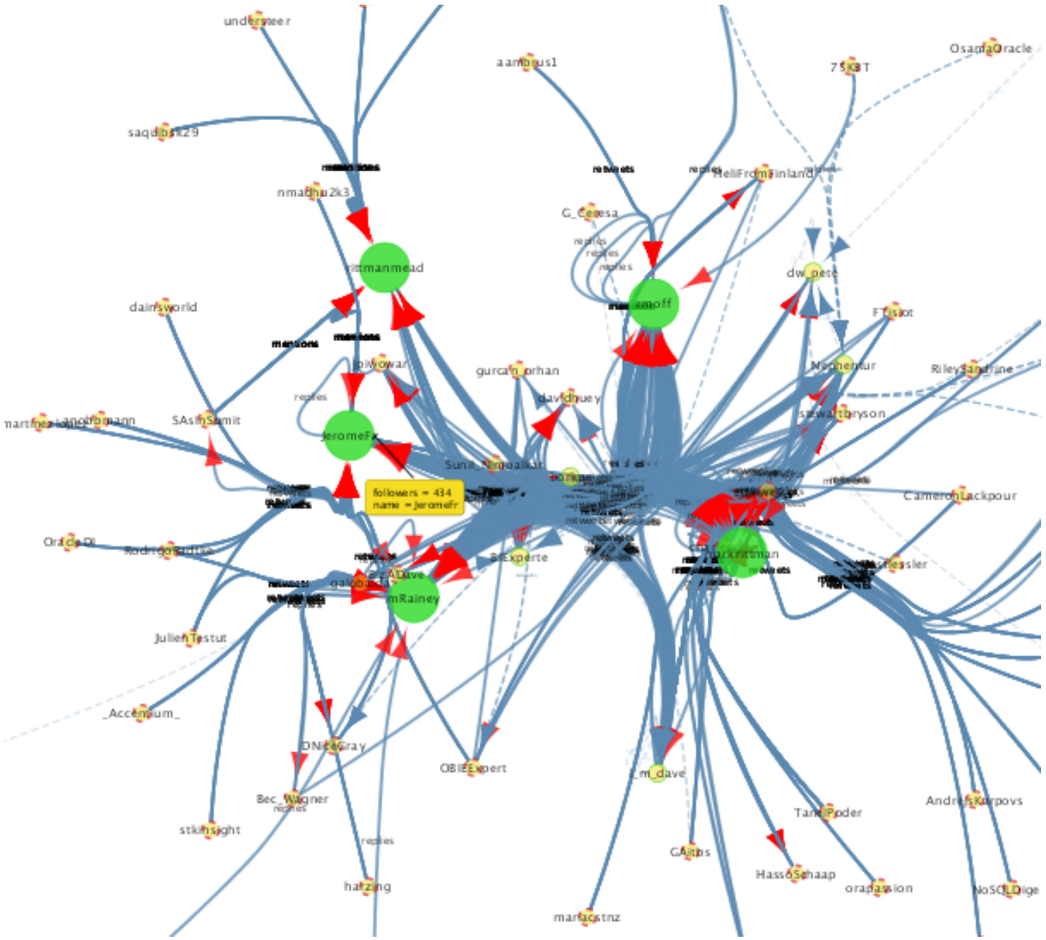


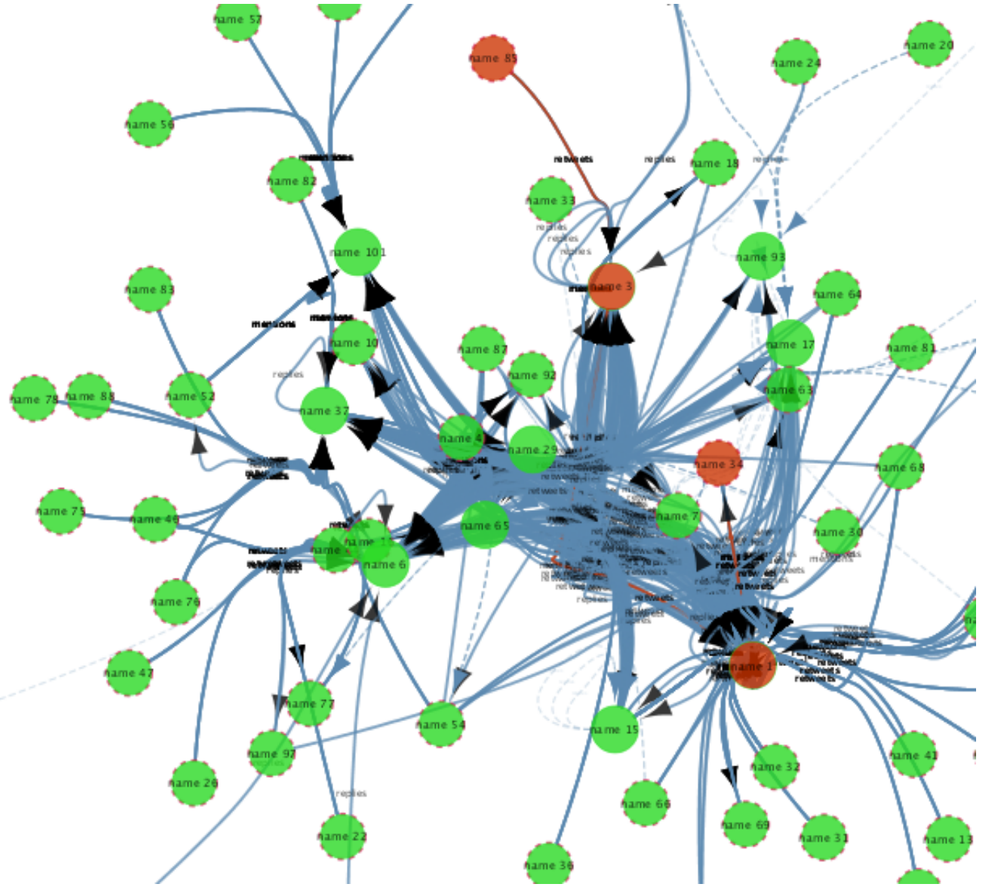
Calculating Shortest Path Between Users



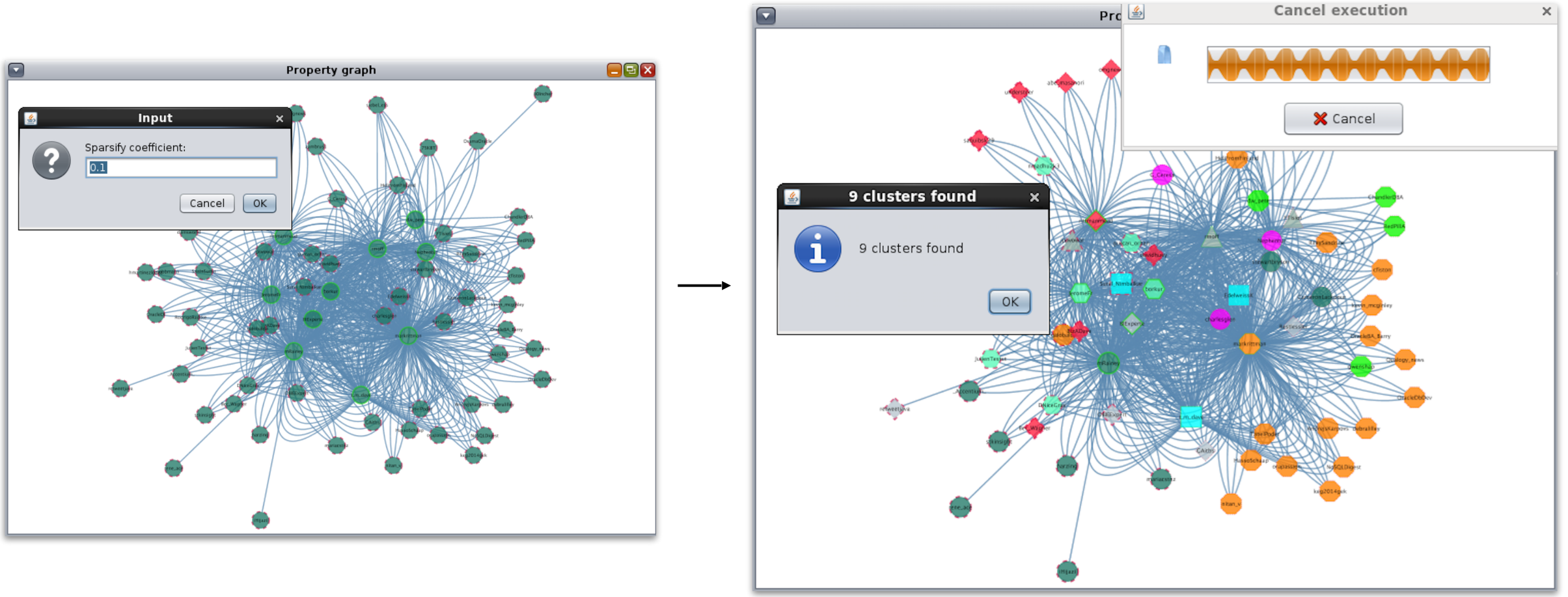


Edge Bundling to Better Illustrate Connection Frequency



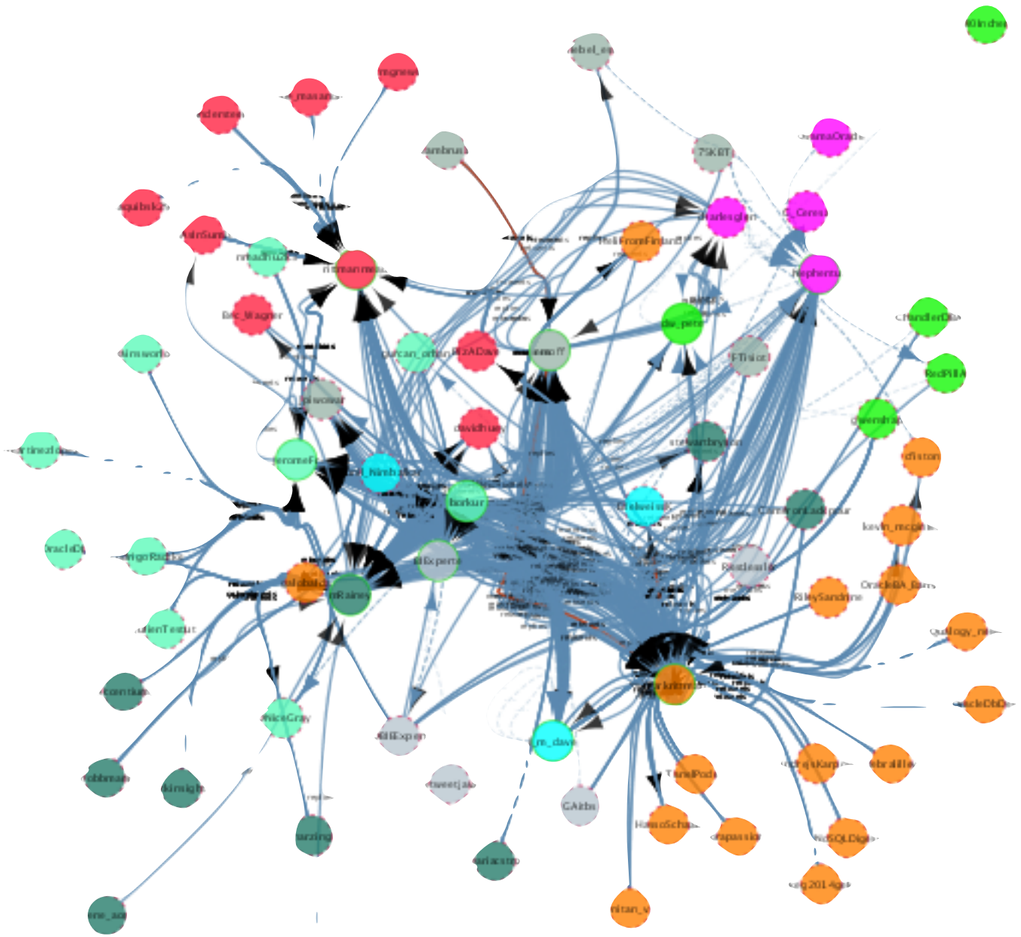


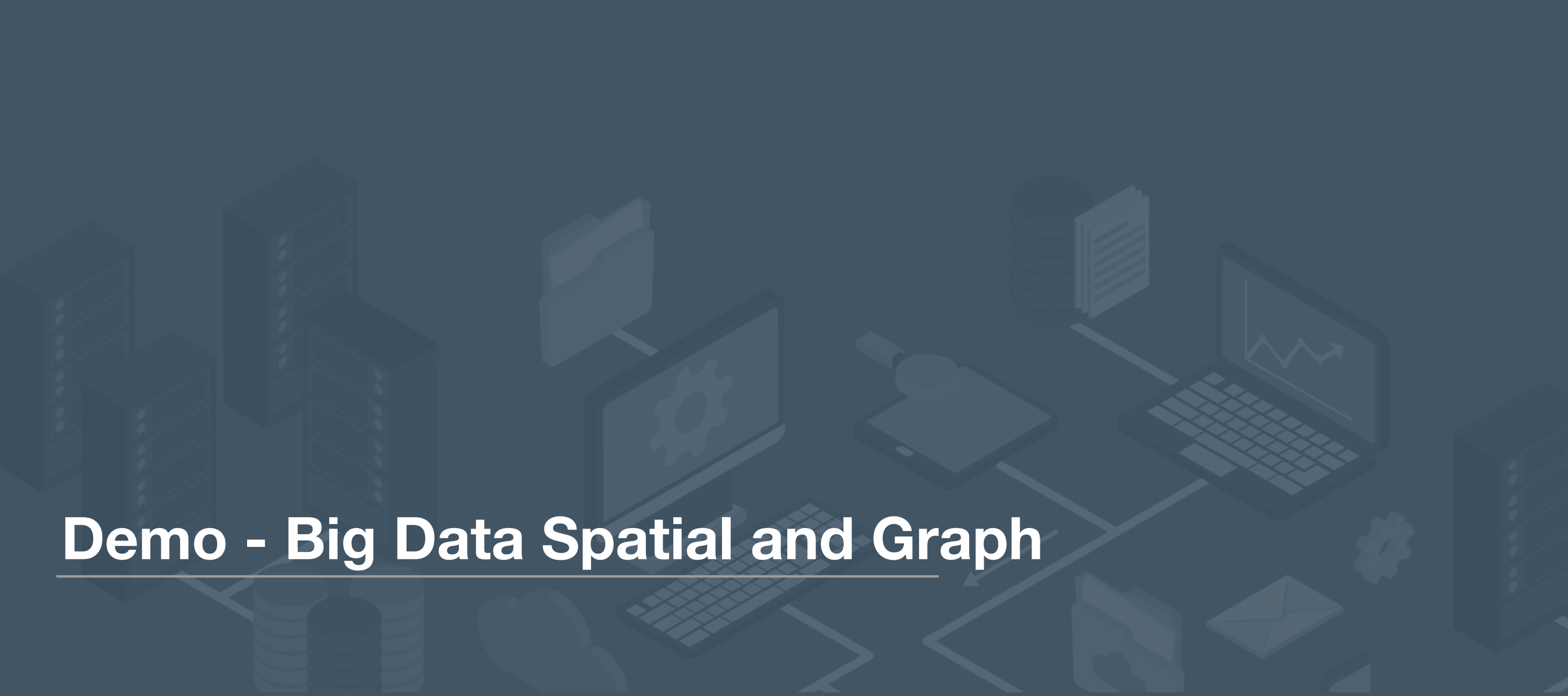
Determining Communities via Twitter Interactions



Determining Communities via Twitter Interactions

- Clusters based on actual interaction patterns, not hashtags
- Detects real communities, not ones that exist just in-theory





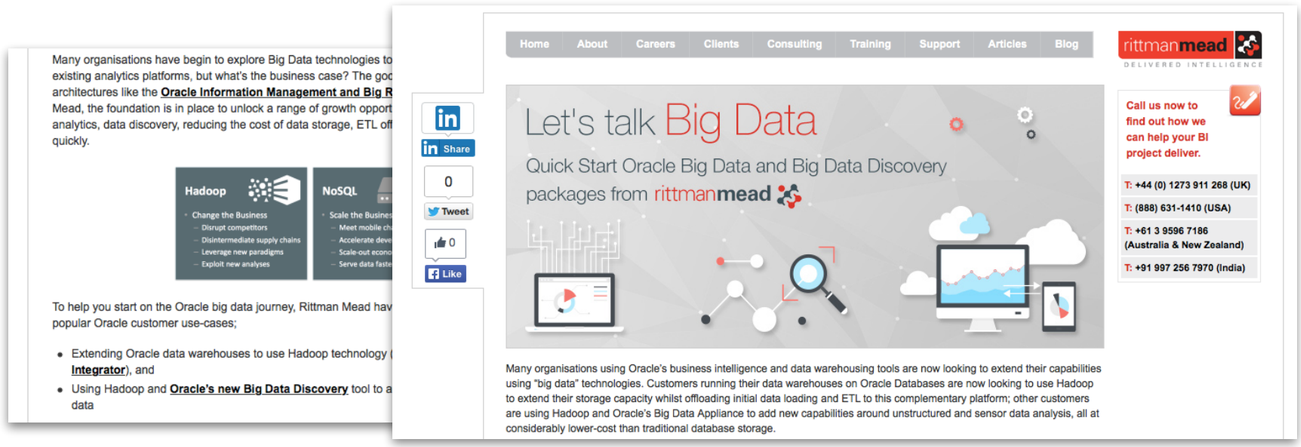
Demo - Big Data Spatial and Graph

Big Data Rapid Start from Rittman Mead

- Extend your organisation's reach into your data with Oracle Big Data Discovery, Cloudera Hadoop and the Rittman Mead Big Data Rapid Start.
- The Big Data Rapid Start is a fixed price, two week engagement delivered by Rittman Mead's team of Oracle, Big Data and Data Discovery consultants, designed to quickly provide everything required to begin discovering the hidden value of your data.
- Move forward with confidence in the technology, process and application of Big Data Discovery with the support of the world's leaders.

Additional Resources

- Articles on the Rittman Mead Blog
  - http://www.rittmanmead.com/category/oracle-big-data-appliance/
  - http://www.rittmanmead.com/category/big-data/
  - http://www.rittmanmead.com/category/oracle-big-data-discovery/
- Rittman Mead offer consulting, training and managed services for Oracle Big Data
  - Oracle & Cloudera partners
  - http://www.rittmanmead.com/bigdata



Unlock the Value in your Big Data Reservoir using Oracle Big Data Discovery and Oracle Big Data Spatial and Graph

Robin Moffatt, Head of R&D (Europe), Rittman Mead

DW & Big Data Global Leaders Program, June 2016 - Oslo, Norway